

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Штучного інтелекту  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

Аналітичний підхід до персоналізованої рекомендації українських казок за допомогою LLM  
(тема)

Виконав:  
здобувач четвертого року навчання,  
групи ІТШ-21-1

Даніела Меллер  
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва спеціальності)

Тип програми освітньо-професійна  
Освітня програма Штучний інтелект  
(повна назва освітньої програми)

Керівник ас. Олександр Стьопін  
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ \_\_\_\_\_  
(підпис)

Олег ЗОЛОТУХІН  
(власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ Комп'ютерних наук \_\_\_\_\_

Кафедра \_\_\_\_\_ Штучного інтелекту \_\_\_\_\_

Рівень вищої освіти \_\_\_\_\_ перший (бакалаврський) \_\_\_\_\_

Спеціальність \_\_\_\_\_ 122 Комп'ютерні науки \_\_\_\_\_  
(код і повна назва)

Тип програми \_\_\_\_\_ освітньо-професійна \_\_\_\_\_

Освітня програма \_\_\_\_\_ Штучний інтелект \_\_\_\_\_  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві \_\_\_\_\_ Меллер Даніелі Михайлівні \_\_\_\_\_  
(прізвище, ім'я, по батькові)

1. Тема роботи \_\_\_\_\_ Аналітичний підхід до персоналізованої рекомендації українських казок за допомогою LLM \_\_\_\_\_

затверджена наказом університету від 19 травня 2025 р. № 265Ст

2. Термін подання студентом роботи до екзаменаційної комісії 20 червня 2025 р.

3. Вихідні дані до роботи \_\_\_\_\_ Науково-технічні публікації, дані Інтернет-джерел та відомих наукових проєктів, Ollama документація, Flask документація \_\_\_\_\_

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

1) Аналіз та постановка задачі \_\_\_\_\_

2) Огляд існуючих рекомендаційних систем казок \_\_\_\_\_

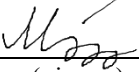
3) Аналітичний аналіз LLM та розробка рекомендаційної системи казок \_\_\_\_\_

4) Розроблена рекомендаційна система казок \_\_\_\_\_

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	19.05.2025	виконано
2	Аналіз предметної галузі	23.05.2025	виконано
3	Постановка задачі	25.05.2025	виконано
4	Огляд існуючих рішень рекомендації казок	29.05.2025	виконано
5	Розробка алгоритму для аналітичного аналізу LLM моделей	30.05.2025	виконано
6	Проведення аналітичного аналізу та опис отриманих результатів	03.06.2025	виконано
7	Практична реалізація застосунку	07.06.2025	виконано
8	Написання пояснювальної записки	10.06.2025	виконано
9	Перевірка на академічний плагіат	12.06.2025	виконано
10	Нормоконтроль	13.06.2025	виконано
11	Підготовка презентації та доповіді	15.06.2025	виконано
12	Попередній захист	17.06.2025	виконано
13	Рецензування	18.06.2025	виконано
14	Захист перед ЕК	19.06.2025	

Дата видачі завдання 19 травня 2025 р.

Здобувач   
(підпис)

Керівник роботи \_\_\_\_\_ ас. Олександр Стьопін  
(підпис) (посада, власне ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка: 66 с., 21 рис., 1 табл., 1 дод., 20 джерел.

КАЗКИ, ВЕЛИКІ МОВНІ МОДЕЛІ, ДИТЯЧА ЛІТЕРАТУРА, КОНТЕНТ-АНАЛІЗ, ОБРОБКА ПРИРОДНОЇ МОВИ, ПОРІВНЯЛЬНИЙ АНАЛІЗ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, FLASK, PYTHON.

Об'єкт дослідження – аналітичний підхід до персоналізованої рекомендації українських казок за допомогою LLM.

Предмет дослідження – дослідження особливостей тексту дитячих казок, методів класифікації та рекомендації казок, різновидів LLM для текстового аналізу, методів оцінки LLM у виконанні завдань з надання рекомендацій.

Мета роботи – аналіз та порівняння LLM моделей у сфері надання рекомендацій дитячої літератури, розробка застосунку для завантаження тексту дитячих казок для їх подальшої обробки LLM моделями та проведення емпіричних досліджень.

Методи дослідження – теоретичний (збір та структуризація теоретичного матеріалу), експериментальний (програмна реалізація веб-додатку). Методи розробки базуються на мовах програмування Python та технологіях Flask, Tensorflow.

У ході виконання кваліфікаційної роботи розроблено застосунок для надання рекомендованих дитячих казок за допомогою засобів LLM, проведено емпіричне дослідження та порівняння різних LLM для поставленої задачі рекомендації казок.

## **ABSTRACT**

Bachelor's thesis contains: 66 pp., 21 fig., 1 tabl., 1 ann., 20 references.

CHILDREN'S LITERATURE, COMPARATIVE ANALYSIS, CONTENT ANALYSIS, FAIRY TALES, FLASK, LARGE LANGUAGE MODELS, NATURAL LANGUAGE PROCESSING, PYTHON, RECOMMENDATION SYSTEMS.

The object of the study is an analytical approach to personalized recommendations of Ukrainian fairy tales using LLM.

The subject of the study is the investigation of the characteristics of children's fairy tale texts, methods of classification and recommendation of fairy tales, types of LLM for text analysis, and methods of evaluating LLM in performing recommendation tasks.

The purpose of the work is to analyze and compare LLM models in the field of children's literature recommendations, develop an application for uploading children's fairy tale texts for further processing by LLM models, and conduct empirical research.

Research methods: theoretical (collection and structuring of theoretical material), experimental (program implementation of a web application). The development methods are based on the Python programming language and Flask and Tensorflow technologies.

During the course of the qualification work, an application was developed to provide recommended children's fairy tales using LLM tools, empirical research was conducted, and various LLM models were compared for the task of recommending fairy tales.

## ЗМІСТ

Вступ.....	7
1 Аналіз та постановка задачі .....	10
1.1 Аналіз предметної області.....	10
1.1.1 Класифікація казок .....	13
1.1.2 LLM у сфері аналізу тексту .....	22
1.2 Постановка задачі.....	27
2 Огляд існуючих рекомендаційних систем казок .....	31
3 Аналітичний аналіз LLM та розробка рекомендаційної системи казок...	39
4 Розроблена рекомендаційна система казок .....	54
Висновки .....	61
Перелік джерел посилання .....	63
Додаток А Відомість кваліфікаційної роботи .....	66

## ВСТУП

Культурне та освітнє значення казок для когнітивного та емоційного розвитку дітей добре задокументоване в науковій літературі. Ці оповіді, що часто характеризуються фантастичними елементами, моральними засадами та архетипними персонажами, збереглися в різних культурах і часових контекстах, слугуючи засобом передачі суспільних цінностей, етичних засад і культурних знань. Сучасні педагогічні дослідження все частіше наголошують на багатогранних перевагах казок в освітньому середовищі, зокрема на їхній здатності покращувати засвоєння мови, стимулювати уяву та сприяти розвитку емоційного інтелекту у молодих учнів.

У сучасному цифровому світі, що характеризується безпрецедентним доступом до величезних сховищ текстового контенту, педагоги, батьки та юні читачі стикаються зі значними труднощами в навігації та виборі відповідних літературних матеріалів, які відповідають конкретним потребам розвитку, інтересам та культурному контексту. Поширення цифрових бібліотек та освітніх платформ загострило цю проблему, створивши те, що деякі дослідники називають «перевантаженням контентом», коли велика кількість доступних матеріалів парадоксальним чином перешкоджає ефективному використанню ресурсів. Це явище зумовлює необхідність розробки складних рекомендаційних систем, здатних аналізувати характеристики контенту та пропонувати релевантні матеріали, що відповідають педагогічним цілям та індивідуальним уподобанням.

Паралельно з розвитком доступності цифрового контенту, значний прогрес у галузі штучного інтелекту, особливо в сфері обробки природної мови (NLP), створив безпрецедентні можливості для автоматизованого текстового аналізу та рекомендацій. Великі мовні моделі (LLM), що представляють сучасну межу в технології NLP, демонструють чудові можливості в розумінні семантичного змісту, визначенні наративних

структур і розпізнаванні тематичних елементів у літературних текстах. Ці можливості позиціонують LLM як потенційно цінні інструменти для вирішення проблем, пов'язаних з рекомендацією дитячої літератури.

Це дослідження розглядає фундаментальне питання про те, наскільки ефективно різні великі мовні моделі можуть аналізувати та класифікувати дитячі казки, щоб генерувати змістовні та відповідні рекомендації. Дослідження ґрунтується на гіпотезі, що LLM з їхнім поглибленим розумінням лінгвістичних моделей і наративних структур можуть виявити суттєві подібності між казками, які виходять за межі поверхневих текстових характеристик і охоплюють тематичні елементи, архетипи персонажів, емоційну валентність та виховну цінність.

Основними завданнями цього дослідження є:

- розробити комплексну веб-систему, яка дозволить користувачам завантажувати тексти казок і отримувати рекомендації на основі контент-аналізу, виконаного кількома LLM;

- впровадити та інтегрувати різні архітектури LLM у систему, створивши основу для порівняльного аналізу їхньої ефективності в конкретній галузі рекомендацій дитячої літератури;

- створити та застосувати суворі метрики оцінювання для визначення точності, релевантності, різноманітності та освітньої придатності рекомендацій, створених різними LLM;

- проаналізувати та задокументувати порівняльні сильні та слабкі сторони обраних LLM в контексті рекомендацій на основі дитячих казок, додавши емпіричні дані про застосування LLM у спеціалізованих галузях;

- визначити оптимальні підходи для впровадження LLM у системи рекомендацій освітнього контенту, що потенційно може стати основою для майбутніх розробок освітніх технологій.

Це дослідження робить внесок у багато перехресних галузей, включаючи освітні технології, комп'ютерну лінгвістику та дитяче

літературознавство. Його значення можна сформулювати в кількох аспектах.

З педагогічної точки зору, розробка складних рекомендаційних систем для дитячої літератури задовольняє критичну потребу в сучасному освітньому середовищі, потенційно підвищуючи точність, з якою освітні ресурси відповідають індивідуальним навчальним потребам. Оскільки навчальні заклади все частіше застосовують персоналізовані підходи до навчання, інструменти, які полегшують вибір ресурсів, стають важливими для підтримки диференційованого навчання.

У галузі комп'ютерної лінгвістики це дослідження надає цінну інформацію про можливості та обмеження сучасних технологій LLM у застосуванні до спеціалізованих текстових жанрів з чіткими лінгвістичними та структурними характеристиками. Дитяча література з її унікальними нарративними конвенціями, спрощеною лексикою та культурними особливостями створює проблеми для систем обробки природної мови, які переважно навчалися на текстах, орієнтованих на дорослих.

З точки зору технологічної реалізації, дослідження демонструє практичні методології інтеграції декількох систем ШІ в цілісну структуру додатків, вирішуючи проблеми, пов'язані з управлінням API, розподілом обчислювальних ресурсів і дизайном користувацького досвіду в додатках зі штучним інтелектом. Ці реалізації можуть слугувати архітектурними моделями для подібних освітніх технологічних ініціатив.

Крім того, порівняльний аналіз ефективності різних LLM надає емпіричні дані для постійних дискусій щодо оцінки та вибору відповідних технологій штучного інтелекту для конкретних галузей і випадків використання, що потенційно може допомогти в процесах прийняття рішень при закупівлі та впровадженні освітніх технологій.

## 1 АНАЛІЗ ТА ПОСТАНОВКА ЗАДАЧІ

### 1.1 Аналіз предметної галузі

Еволюція казки як окремого літературного жанру являє собою складну траєкторію, що перетинає усні традиції, літературні адаптації та сучасні реінтерпретації. Науковий дискурс про походження казки переважно визнає її появу з фольклорних усних наративів, що передавалися з покоління в покоління, з подальшою літературною «кодіфікацією», що відбулася завдяки цілеспрямованим зусиллям істориків культури та літературних діячів зі збору та транскрипції. Цей перехід від усної до текстової маніфестації становить ключову трансформацію в онтологічному статусі казки, завдяки чому раніше плинний наратив стабілізувався в канонічних формах, які продовжують чинити значний вплив на сучасні концептуалізації жанру.

Фундаментальний внесок Шарля Перро у Франції кінця XVII століття, коли він опублікував «Історії або розповіді про минулі часи», ознаменував формотворчий момент у літературній формалізації казки в європейських культурних контекстах. Адаптація Перро усних народних оповідань для аристократичної аудиторії започаткувала процес удосконалення наративу, що характеризувався введенням чітких моральних настанов та стилістичних прикрас, які відповідали тогочасним літературним конвенціям. Це редакторське втручання створило прецедент для наступних збірок казок, де трансформація фольклору в літературні тексти спричинила значні кураторські рішення щодо відповідності змісту, наративної структури та ідеологічної спрямованості.

Брати Грїмм, чия збірка «Діти та домашні тварини» є чи не найвпливовішою збіркою казок у західній літературній традиції, стали ще одним прикладом цього процесу літературної адаптації. Методологія збірки, яка спочатку наголошувала на філологічній точності в

документуванні германських народних наративів, еволюціонувала в наступних виданнях у бік більшої літературної витонченості та педагогічної придатності для сімейного читання середнього класу. Поступове вилучення зі збірки братів Грімм відвертого насильства, сексуального змісту та антиавторитарних елементів ілюструє гнучкість казок як культурних артефактів, що реагують на мінливі суспільні норми щодо дитинства, освіти та моральної поведінки.

Транскультурна дифузія казкових мотивів становить ще один важливий вимір їхнього історичного розвитку. Порівняльний аналіз наративних моделей у географічно розрізненіх культурах виявив дивовижну схожість у фундаментальних сюжетних структурах, архетипах персонажів і тематичних елементах.

На рисунку 1.1 наведено популярні казки за регіонами світу.



Рисунок 1.1 – Популярні казки за регіонами світу

Монументальний «Індекс мотивів народної літератури» Томпсона та система класифікації типів казок Аарне і Томпсона задокументували ці міжкультурні зближення, що спричинило наукові дебати щодо механізмів передачі наративів. Теорії, що стверджують пряму культурну передачу, конкурують з поясненнями, які приписують схожість спільним психологічним структурам або універсальному людському досвіду.

Дев'ятнадцяте та початок двадцятого століть стали свідками дедалі більшої інституціоналізації казки в дитячій літературі – процесу, якому сприяли нові концепції дитинства як особливої стадії розвитку, що потребує спеціалізованих навчальних матеріалів. Таке педагогічне привласнення казок збіглося з їхнім естетичним піднесенням завдяки ілюстрованим виданням видатних художників, зокрема Артура Рекхема, Едмунда Дулака та Кея Нільсена, чий візуальні інтерпретації суттєво вплинули на культурну уяву щодо естетики казки.

Яскравим прикладом творчості Едмунда Дулака поміж безлічі його робіт можна вважати ілюстрацію до казки «Райський сад», яка зображена на рисунку 1.2.

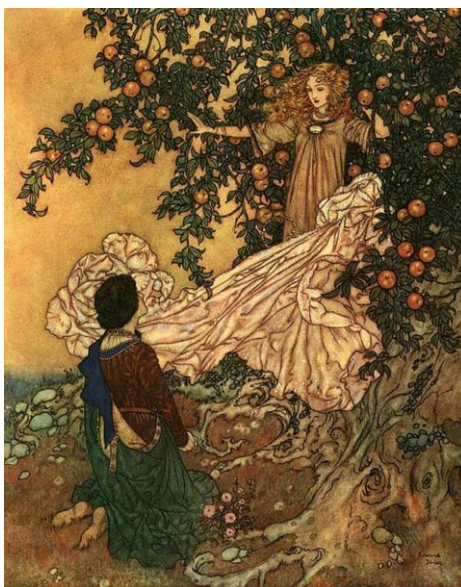


Рисунок 1.2 – Ілюстрація Едмунда Дулака до казки «Райський сад»

Таким чином, ілюстрована книга казок стала особливим матеріальним артефактом, що опосередковує залучення дитини до наративного змісту через мультимодальний досвід читання.

Трансформації цифрових медіа у двадцять першому столітті запровадили нові способи передачі та сприйняття казок, зокрема інтерактивні наративи, кінематографічні адаптації та трансмедійні підходи до оповіді. Ці технологічні розробки сприяли появі інноваційних можливостей, водночас піднімаючи питання щодо авторитету тексту, канонічної стабільності та меж загальної класифікації в епоху медіаконвергенції. Живучість казки під час цих технологічних змін свідчить про її неабияку культурну стійкість та адаптаційну здатність.

### 1.1.1 Класифікація казок

Класифікація казок становить значні методологічні виклики через їхню родову гібридність, культурну варіативність та історичну мінливість. Тим не менш, було розроблено кілька впливових класифікаційних систем, які забезпечують аналітичні рамки для виявлення відмінних рис та організації наративних варіантів у цілісні типологічні схеми.

Система класифікації Аарне-Томпсона-Утера (ATU) представляє найповніший і найпоширеніший таксономічний підхід до народних наративів, включно з казками. Ця система класифікує наративи за типами казок, які являють собою впізнавані сюжетні структури, що зберігаються в різних переказах, незважаючи на варіації в окремих деталях. В індексі ATU цим типам казок присвоюються цифрові позначення, причому номери 300-749 спеціально призначені для «Казок про магію» – категорії, яка охоплює те, що в популярному дискурсі зазвичай ідентифікується як просто казки. Ця система класифікації виявилася безцінною для порівняльного аналізу наративних традицій у різних культурних контекстах, полегшуючи виявлення транснаціональних моделей і відмінних регіональних варіантів.

Морфологічний аналіз Проппа, вперше опублікований в 1928 році, пропонує альтернативний підхід до класифікації казок, зосереджений на наративній функції, а не на тематичному змісті. Систематично досліджуючи народні казки, Пропп виділив тридцять одну функцію, що представляють фундаментальні дії, які структурують ці казки, стверджуючи, що ці функції послідовно з'являються в інваріантному послідовному порядку, незважаючи на значну варіативність атрибутів персонажів і конкретних деталей сюжету. Цей підхід суттєво вплинув на подальший аналіз казок, надавши методологічні інструменти для вивчення їхніх композиційних моделей незалежно від поверхневих тематичних елементів.

За допомогою функцій Проппа кожен казку можна розбити на окремі частини, а потім зібрати подібно до конструктору, як зображено на рисунку 1.3.

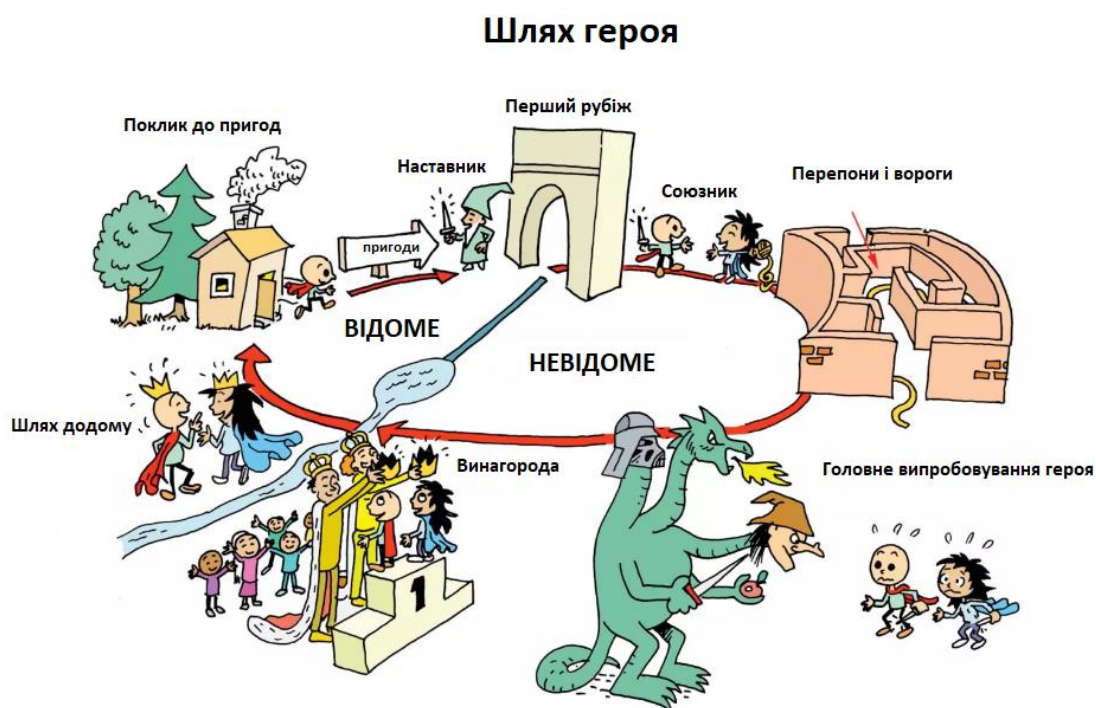


Рисунок 1.3 – Приклад казки, створений за допомогою функцій Проппа

Феноменологічний підхід Люті до казок визначає кілька відмінних стилістичних і формальних характеристик, які відрізняють їх від інших

фольклорних жанрів. До них належать одновимірність (безшовна інтеграція магічних і мирських елементів), безглибинність (персонажі позбавлені психологічної складності), абстрактний стиль (мінімальна описова деталізація), ізольованість і універсальний взаємозв'язок (протагоністи відірвані від соціальних мереж, але пов'язані з різноманітними наративними елементами), а також сублимація і всеохопність (перетворення реалістичних елементів на абстрактні наративні компоненти). Формалістичний акцент Люті на цих відмінних естетичних якостях доповнює структурний аналіз, запропонований Проппом, і тематичну класифікацію системи АТУ.

Підходи до класифікації на основі змісту визначили кілька повторюваних тематичних елементів, характерних для казок, зокрема наявність чарівних предметів або надприродного втручання, трансформаційні подорожі, які здійснюють герої, конфлікти між чітко окресленими моральними силами та сюжетні розв'язки, які зазвичай відновлюють соціальний порядок завдяки підвищенню статусу героя або його багатству. Ці тематичні патерни часто проявляються у впізнаваних наративних сюжетах, таких як переслідуваний герой, чарівні помічники, нездійсненні завдання та трансформаційні сцени впізнавання, які структурують розвиток від початкової рівноваги через порушення до остаточної розв'язки.

Лінгвістичні особливості казок становлять ще один вимір їхньої родової відмінності. Стилістичні маркери включають формульні початкові та кінцеві фрази («Жили-були», «Жили вони довго і щасливо»), паратактичний синтаксис, що підкреслює послідовну дію, а не причинно-наслідкові зв'язки, триразове повторення подій або дій з прогресуючою інтенсифікацією, а також мінімальну психологічну експозицію. Ці лінгвістичні характеристики створюють особливий наративний голос, що асоціюється з казкою, який характеризується всезнаючою перспективою, що представляє події з мінімальним емоційним забарвленням або моральною оцінкою.

Цифрові підходи нещодавно запровадили обчислювальні методи для класифікації казок, використовуючи методи обробки природної мови для виявлення лінгвістичних моделей, нарративних структур і тематичних елементів у великих оцифрованих текстах. Ці кількісні підходи доповнюють традиційний якісний аналіз, виявляючи закономірності, які можуть бути не одразу помітні за допомогою методів уважного читання, хоча вони обов'язково спираються на попередню розробку надійних схем класифікації для інформування алгоритмічного дизайну.

Ранні обчислювальні методології, як правило, використовували статистичні підходи до аналізу текстів, застосовуючи такі методи, як частота терміна, обернена до частоти документа (TF-IDF), для виявлення характерних лексичних моделей і тематичних маркерів у текстах.

Приклад визначення показника TF-IDF наведений на рисунку 1.4.

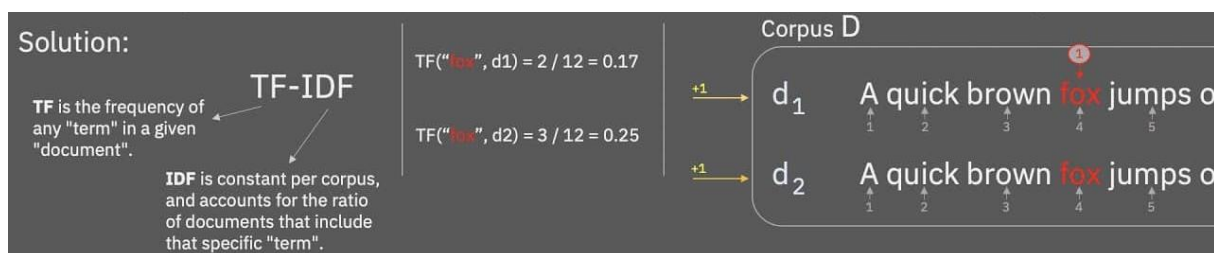


Рисунок 1.4 – Приклад визначення показника TF-IDF

Ці лексичні підходи надавали кількісні показники схожості текстів на основі моделей розподілу лексики, хоча їм часто було важко охопити семантичні зв'язки та нарративні структури, які виходять за рамки аналізу на рівні слів.

Згодом підходи машинного навчання вдосконалили комп'ютерну категоризацію казок завдяки застосуванню алгоритмів керованого та некерованого навчання до казок. Методи кластерного аналізу визначили групи казок на основі декількох текстових ознак, виявивши патерни схожості, які можуть бути не одразу помітними при ручній класифікації.

Алгоритми класифікації, навчені на колекціях казок, класифікованих вручну, продемонстрували здатність поширювати схеми категоризації на раніше некатегоризовані тексти на основі вивчених патернів ознак, хоча ефективність цих підходів залишається залежною від якості та повноти навчальних даних.

Підходи тематичного моделювання, зокрема латентний розподіл Діріхле (LDA), були застосовані до колекцій казок для виявлення тематичних патернів, розподілених у текстах.

Опис алгоритму LDA наведено на рисунку 1.5.

Setup	Collecting Text documents
Preprocessing	Tokenization, Removing Stop words, Stemming Vocabulary creation
Document Matrix	Creation of document-term matrix, where rows= documents & column = term
Model Initialization	Define: No. of topics to discover(hyperparameter) = $K$
Dirichlet Distribution	Assign: Prior for document-topic distribution: $\alpha$ Prior for topic-word distribution: $\beta$
Document Matrix	Creation of document-term matrix, where rows= documents & column = term
Topic Initialization	For each word in each document: Randomly assign topic from 1 to $K$ .
Gibbs Sampling	Calculate and sample a new topic assignment based on dirichlet & update topic assignment for words
Parameter estimation	Document-Topic in each document: $\vartheta$ Topic- word in each document: $\phi$

Рисунок 1.5 – Опис алгоритму LDA

Ці методи неконтрольованого навчання визначають імовірнісний розподіл тем у документах, що дає змогу виявити тематичні спільності, які

можуть стати основою для рекомендаційних рішень. Застосоване до казок, моделювання виявило тематичні кластери, пов'язані з сімейними стосунками, надприродними елементами, моральними конфліктами та соціальною мобільністю, які забезпечують багатовимірні рамки для оцінки схожості, що виходять за межі традиційних класифікаційних категорій.

Нещодавні розробки в галузі семантичного аналізу застосували методи вбудовування слів, зокрема Word2Vec і GloVe, до текстів казок, що уможливило векторне представлення семантичних зв'язків, які фіксують концептуальну схожість, що виходить за межі лексичного збігу.

Приклад візуалізації слів за допомогою Word2Vec наведено на рисунку 1.6.

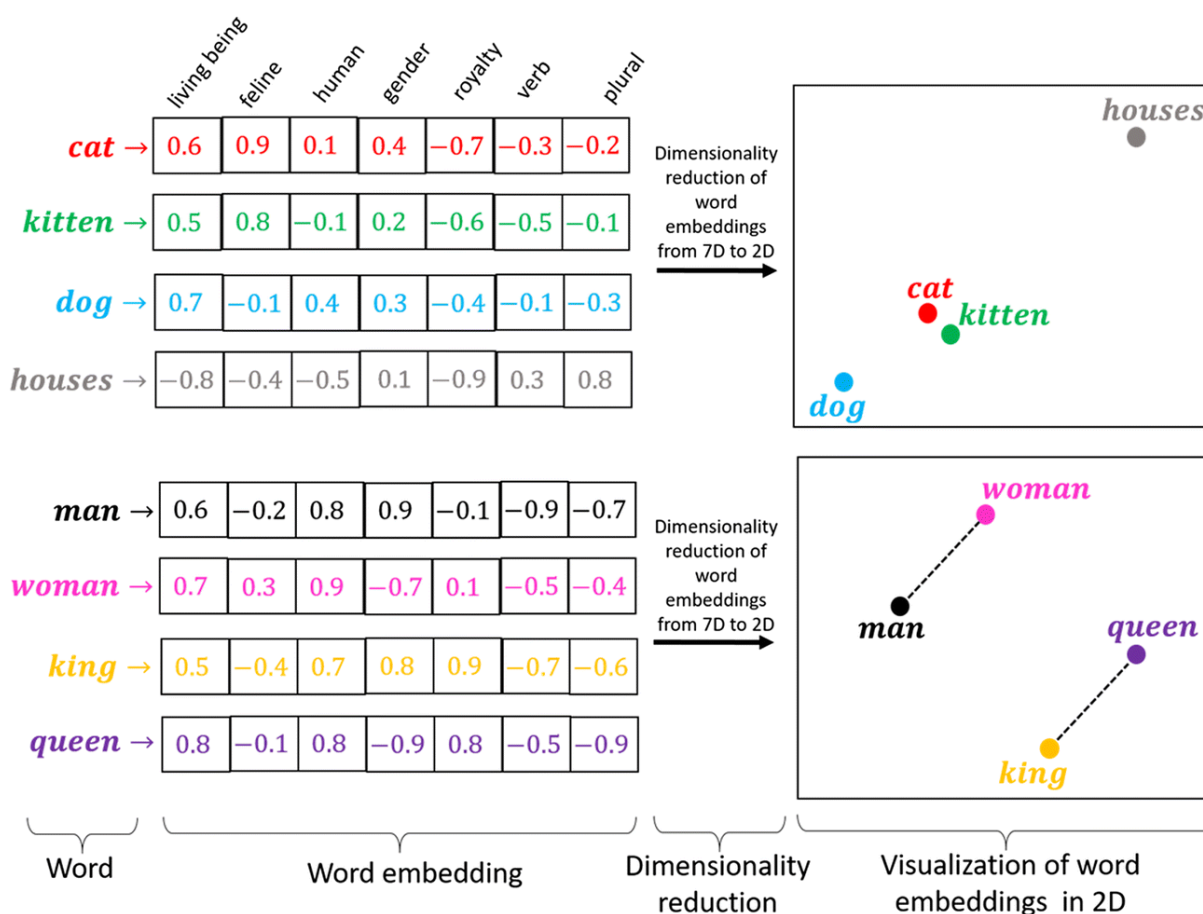


Рисунок 1.6 – Приклад застосування Word2Vec

Ці підходи відображають слова у високорозмірні векторні простори, де семантична близькість представлена через просторові відношення, що дозволяє проводити більш складні вимірювання подібності, які можуть виявити концептуальні паралелі між текстами, незважаючи на поверхневі відмінності в лексиці. У застосуванні до казок ці семантичні підходи продемонстрували здатність виявляти тематичні та концептуальні подібності, які узгоджуються з людськими судженнями щодо нарративних зв'язків.

Методи аналізу настроїв були застосовані до категоризації казок для виявлення емоційних патернів у наративах, надаючи додаткові параметри для оцінки схожості та рекомендацій. Ці підходи аналізують емоційну валентність лексики, структури речень і розвиток оповіді, щоб створити профілі емоційного змісту, які можуть бути використані для визначення вікової придатності або прогнозування емоційного резонансу. У застосуванні до казок аналіз настроїв виявив особливі емоційні траєкторії, що характеризуються вирішенням негативних емоційних станів через розвиток оповіді, надаючи потенційні індикатори для рекомендацій, заснованих на вирівнюванні емоційного змісту.

Аналіз нарративної структури є особливо перспективним напрямком для комп'ютерної категоризації казок, оскільки він намагається виявити структурні патерни, які виходять за рамки особливостей на рівні слів або речень і охоплюють загальну нарративну організацію. Комп'ютерні адаптації морфологічного підходу Проппа намагалися автоматично ідентифікувати нарративні функції в текстах, уможливлуючи структурне порівняння на основі патернів розвитку оповіді, а не поверхневої схожості змісту. Ці підходи залишаються складними для ефективного застосування через складність виявлення нарративної структури, хоча нещодавні досягнення в розумінні природної мови продемонстрували зростаючу спроможність ідентифікувати нарративні компоненти, стосунки між персонажами та

сюжетні структури, які можуть бути використані для оцінки структурної подібності.

Поява мовних моделей на основі трансформерів, зокрема BERT (Bidirectional Encoder Representations from Transformers) та її похідних, відкрила нові можливості для аналізу казок за допомогою контекстуальних вбудовувань, які фіксують семантичні зв'язки з більшою кількістю нюансів, ніж попередні підходи до вбудовування слів. Ці моделі, які включають механізми уваги та двонаправлені підходи до навчання, демонструють покращену здатність розуміти контекстуальне значення, основні зв'язки та семантичні структури, що уможливорює більш складний аналіз тексту. У застосуванні до казок ці моделі потенційно пропонують покращену здатність виявляти тематичні паралелі, стосунки між персонажами та наративні структури, які можуть впливати на прийняття рекомендаційних рішень.

Сучасні системи рекомендацій для дитячої літератури дедалі частіше застосовують гібридні підходи, які поєднують кілька аналітичних вимірів, зокрема контент-аналіз, дані про поведінку користувачів і контекстуальні чинники. Методи колаборативної фільтрації виявляють закономірності в даних про залучення користувачів, щоб генерувати рекомендації на основі схожості вподобань користувачів, а не текстового контенту, доповнюючи підходи, засновані на контенті, поведінковими уявленнями. Системи освітніх рекомендацій включають педагогічні параметри, зокрема рівень читання, освітні цілі та відповідність рівню розвитку як додаткові фактори у формуванні рекомендацій, намагаючись узгодити літературні добірки з освітнім контекстом та навчальними цілями.

Незважаючи на цей технологічний прогрес, у розробці системи рекомендацій казок залишаються значні проблеми. Серед них – обмежена доступність повних оцифрованих колекцій казок з узгодженим форматуванням і метаданими, труднощі з фіксацією культурних нюансів і контекстуальних знань, які формують людську інтерпретацію цих текстів, а

також інтеграція міркувань вікової придатності, які виходять за рамки комп'ютерного контент-аналізу і включають в себе знання з психології розвитку. Крім того, оцінка якості рекомендацій створює методологічні проблеми, оскільки суб'єктивні оцінки схожості наративів і доречності рекомендацій можуть значно відрізнятись серед користувачів і в різних культурних контекстах.

Застосування великих мовних моделей для аналізу та рекомендацій казок є новим етапом у цій галузі, що потенційно долає деякі обмеження попередніх обчислювальних підходів завдяки покращеному семантичному розумінню, контекстуальній обізнаності та можливостям розуміння наративу. Ці моделі, навчені на великих текстових масивах за допомогою підходів самонавчання, демонструють безпрецедентну здатність розуміти лінгвістичні нюанси, тематичний зміст і наративну структуру завдяки тому, що вони стикаються з різноманітними текстовими патернами. Їхнє потенційне застосування для рекомендацій казок відкриває можливості для більш складних оцінок подібності на основі багатовимірного розуміння наративних характеристик, хоча емпірична оцінка їхньої ефективності в цій специфічній сфері залишається обмеженою в сучасній літературі.

Порівняльна оцінка різних архітектур LLM у завданнях рекомендацій казок становить значну прогалину в сучасній літературі. Хоча загальні оцінки ефективності LLM у різних завданнях обробки природної мови були проведені, їхнє конкретне застосування для рекомендацій дитячої літератури залишається недостатньо вивченим, особливо щодо їхньої відносної ефективності у виявленні значущих подібностей між казками, які узгоджуються з людськими судженнями та освітніми цілями. Ця прогалина підкреслює важливість систематичного порівняльного аналізу для визначення того, які модельні архітектури, навчальні підходи та методології впровадження найбільш ефективно підтримують рекомендації щодо казок в освітніх контекстах.

### 1.1.2 LLM у сфері аналізу тексту

Траєкторія розвитку обробки природної мови (NLP) являє собою вражаючий прогрес від рудиментарних систем, заснованих на правилах, до складних нейронних архітектур, здатних до лінгвістичного розуміння. Ця еволюція відображає як зміну концептуальної парадигми, так і технологічний прогрес, що докорінно змінив обчислювальні підходи до аналізу мови.

Великі мовні моделі – це сучасна кульмінація технології обробки природної мови, що характеризується безпрецедентним масштабом, архітектурною складністю та неабиякими лінгвістичними можливостями. Розуміння їхніх функціональних механізмів вимагає вивчення їхніх фундаментальних архітектурних компонентів, методологій навчання та принципів роботи.

Архітектура-трансформер формує фундаментальну структуру сучасних LLM, що вирізняється своїми механізмами самоуваги, які встановлюють взаємозв'язки між усіма позиціями у вхідній послідовності.

Структуру трансформера наведено на рисунку 1.7.

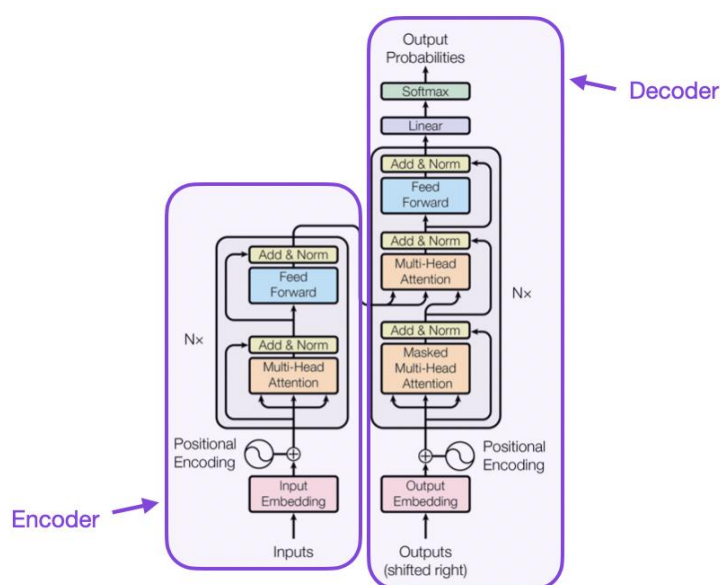


Рисунок 1.7 – Структура трансформера

Базовий блок трансформера складається з шару самоуваги, за яким слідує позиційна мережа прямого поширення, із залишковими зв'язками і нормалізацією шарів, що застосовуються для забезпечення стабільності навчання. Така структура забезпечує ефективну паралельну обробку, зберігаючи при цьому здатність моделювати складні залежності між віддаленими позиціями у вхідній послідовності.

Механізми самоуваги становлять основну інновацію трансформера, обчислюючи ваги уваги, які визначають, як кожна позиція повинна приділяти увагу всім іншим позиціям при побудові свого контекстного представлення. Багатоголова реалізація цього механізму дозволяє моделі одночасно обробляти інформацію з різних підпросторів репрезентації, фіксуючи різноманітні зв'язки між лексемами, включаючи синтаксичні залежності, зв'язки між референтами та семантичні асоціації. Ця функціональність забезпечує архітектурну основу для можливостей контекстного розуміння, які демонструють LLM.

Стратегії позиційного кодування спрямовані на притаманну трансформера позиційну діагностику шляхом введення позиційної інформації в представлення токенів. Ці кодування, реалізовані за допомогою синусоїдальних функцій або вбудовування вивчених позицій, гарантують, що модель може розрізнити входження однієї і тієї ж лексеми в різних позиціях, зберігаючи послідовну інформацію, незважаючи на розпаралелений підхід архітектури до обробки. Ця можливість є важливою для завдань, що вимагають чутливості до порядку слів і послідовної структури.

Сучасні LLM зазвичай реалізують архітектури лише кодера (наприклад, BERT), лише декодера (наприклад, GPT) або архітектури кодера-декодера (наприклад, T5), кожна з яких оптимізована для різних профілів застосування. Моделі, що використовують лише кодер, чудово справляються із завданнями, що вимагають двонаправленого аналізу контексту, включаючи класифікацію та вилучення інформації. Моделі, що

працюють лише з декодером, спеціалізуються на генеративних завданнях завдяки односпрямованій обробці, яка зберігає властивість авторегресії, необхідну для генерації тексту. Моделі кодер-декодер поєднують ці можливості для підтримки послідовних завдань, включаючи узагальнення та переклад.

Масштабування властивостей архітектур трансформерів виявилось напрочуд послідовним: покращення продуктивності тісно корелює зі збільшенням розміру (параметрів) моделі, обсягу навчальних даних та обчислювальних ресурсів, виділених для навчання. Цей взаємозв'язок масштабування, емпірично продемонстрований дослідниками, зокрема Капланом та Брауном, мотивував розробку все більших моделей, і сучасні системи містять сотні мільярдів параметрів, навчених на багатотрильйонних наборах даних токенів.

Цілі попереднього навчання формують фундаментальні можливості LLM, визначаючи навчальні завдання для самоконтролю, які використовуються під час початкового навчання. Моделювання замаскованої мови (передбачення замаскованих лексем з двонаправленого контексту) та каузальне моделювання мови (передбачення наступної лексеми на основі попередніх лексем) є основними завданнями, що використовуються в сучасних моделях, зі спеціалізованими варіантами, що включають маскування діапазону, передбачення порядку речень та виявлення заміненних лексем, які надають додаткові навчальні сигнали. Ці цілі дозволяють моделям розробляти надійні лінгвістичні репрезентації, не вимагаючи вручну анотованих навчальних даних.

Процес навчання для сучасних LLM передбачає багатоетапний підхід, який зазвичай включає попереднє навчання на різноманітних текстових корпусах, за яким слідує спеціалізовані етапи тонкого налаштування або адаптації. На початковому етапі підготовки використовуються великі масиви даних, зібрані з інтернет-текстів, книг, наукових статей та інших джерел, для розвитку загальних лінгвістичних здібностей за допомогою

самоконтрольованих навчальних завдань. Цей етап вимагає значних обчислювальних ресурсів, часто передбачає розподілене навчання між сотнями або тисячами акселераторів протягом тривалих періодів.

Стратегії тонкого налаштування адаптують попередньо навчені моделі до конкретних завдань або доменів шляхом постійного навчання на спеціалізованих наборах даних з конкретними цілями. Цей підхід до навчання з перенесенням використовує надійні лінгвістичні репрезентації, розроблені під час попереднього навчання, одночасно спеціалізуючи модель для конкретних застосувань. Для зменшення обчислювальних витрат, пов'язаних з адаптацією великих моделей до різноманітних завдань, з'явилися такі методи, як точне налаштування з урахуванням параметрів та швидке налаштування.

Налаштування інструкцій, що представляє собою спеціалізовану форму точного налаштування, адаптує моделі до виконання інструкцій природною мовою шляхом навчання на наборах даних, що складаються з пар «інструкція-відповідь». Цей підхід розширює можливості моделей для навчання з нульовою та малою кількістю спроб, розвиваючи загальну здатність слідувати інструкціям, яка переноситься через межі завдань. Ця методологія виявилася особливо цінною для розробки моделей, здатних виконувати різноманітні завдання без тонкого налаштування під конкретне завдання.

Навчання з підкріпленням на основі зворотного зв'язку з людиною (RLHF) є наступним етапом доопрацювання, на якому моделі оптимізуються для узгодження з людськими вподобаннями за допомогою моделювання винагороди на основі людських оцінок результатів моделювання. Цей підхід усуває обмеження в контрольованому доопрацюванні, надаючи механізми для включення суб'єктивних критеріїв якості, зокрема корисності, нешкідливості та чесності, які важко визначити за допомогою звичайних функцій втрат. RLHF стає все більш центральним

елементом у розвитку LLM, оскільки фокус змістився з «сирих» можливостей на узгодження з людськими цінностями та очікуваннями.

Нові можливості представляють особливо важливий аспект функціональності LLM, коли моделі демонструють здібності, які не були явно націлені під час навчання, після досягнення достатнього масштабу. Ці можливості, включаючи контекстне навчання (навчання з кількох спроб без оновлення параметрів), міркування через покрокове пояснення і складні інструкції, виникають як емерджентні властивості масштабу, а не як явно спроектовані функції. Цей феномен свідчить про те, що сучасні моделі можуть являти собою ранні приклади законів масштабування, повні наслідки яких залишаються не до кінця зрозумілими.

Внутрішня репрезентативна динаміка LLMs була досліджена за допомогою механістичного дослідження інтерпретованості, яке виявило диференційовану функціональність на різних рівнях мережі. Нижчі шари зазвичай фіксують лексичні та синтаксичні особливості, середні шари обробляють семантичні зв'язки та композиційне значення, а вищі шари інтегрують ширшу контекстну інформацію та репрезентації для конкретних завдань. Ця функціональна спеціалізація виникає природним чином під час навчання без явної архітектурної диференціації, представляючи собою емерджентну організацію, яка полегшує лінгвістичну обробку моделі.

Представлення знань в LLM відбувається переважно через параметричне кодування у вагах моделі, при цьому фактична інформація, лінгвістичні патерни та процедурні знання розподіляються за параметрами мережі, а не зберігаються у явних декларативних структурах. Таке розподілене представлення дає змогу гнучко застосовувати знання, але створює проблеми з верифікацією, оновленням та атрибуцією знань. Нещодавні підходи, включаючи розширення пошуку та редагування знань, долають ці обмеження, доповнюючи параметричні знання явним пошуком або цілеспрямованою зміною параметрів.

Сучасні архітектури LLM включають різні вдосконалення базової структури трансформера, включаючи ефективні варіанти уваги, які зменшують обчислювальну складність, підходи на основі суміші експертів, які збільшують кількість параметрів без пропорційного збільшення обчислень, та архітектурні модифікації для підтримки довших контекстних вікон. Ці інновації спрямовані на подолання специфічних обмежень стандартних реалізацій трансформерів, зберігаючи при цьому їхні фундаментальні можливості для контекстного представлення та генерації.

Архітектурна складність і масштаб сучасних LLM дозволили створити системи з безпрецедентними можливостями природної мови, включаючи контекстне розуміння, застосування знань і комплексне генерування, які наближаються до людського виконання різноманітних завдань. Хоча сучасні моделі зберігають значні обмеження, їхня траєкторія розвитку передбачає подальший прогрес завдяки подальшому масштабуванню, вдосконаленню архітектури та розробці методології навчання, що матиме суттєві наслідки для додатків, включаючи системи рекомендацій казок, які є основною темою цього дослідження.

## 1.2 Постановка задачі

Це дослідження реалізує веб-систему рекомендацій казок, що використовує декілька LLM для генерування та порівняння рекомендацій на основі завантаженого користувачем контенту. Система використовує клієнт-серверну архітектуру з Flask як веб-фреймворком і Python як основною мовою програмування. Архітектура складається з чотирьох основних компонентів: рівень інтерфейсу користувача, прикладний рівень, механізм рекомендацій та рівень зберігання даних.

Механізм рекомендацій слугує основним компонентом системи, реалізуючи паралельну обробку завантажених користувачем казок за допомогою декількох LLM. Кожен LLM обробляє один і той самий вхідний

текст незалежно, застосовуючи свої унікальні можливості розуміння мови для створення рекомендацій на основі контенту. Така паралельна архітектура дає змогу безпосередньо порівнювати ефективність рекомендацій різних LLM, зберігаючи при цьому однакові вхідні умови в усіх моделях.

Потік даних в системі слідує послідовному конвеєру обробки, починаючи з введення даних користувачем і закінчуючи представленням рекомендацій. Послідовність первинного потоку даних виглядає наступним чином:

- користувач завантажує текст казки через веб-інтерфейс;
- прикладний рівень обробляє та перевіряє завантажений документ;
- модуль вилучення тексту перетворює документ у звичайний текстовий формат;
- модуль попередньої обробки застосовує операції нормалізації, токенизації та очищення;
- механізм рекомендацій розподіляє оброблений текст між декількома екземплярами LLM;
- кожна LLM аналізує вміст і генерує оцінки схожості для записів у базі даних;
- агрегатор результатів збирає та форматує рекомендації від кожного LLM;
- рівень презентації відображає порівняльні результати в інтерфейсі користувача.

Система має асинхронну обробку, яка реалізована для взаємодії LLM, щоб запобігти таймаутам з'єднання під час тривалих операцій обробки. Система використовує чергу повідомлень для управління запитами на обробку LLM, що забезпечує паралельне виконання, зберігаючи при цьому можливість відстеження запитів і обробки помилок.

На кожному етапі обробки збираються показники продуктивності, щоб полегшити подальше порівняння ефективності та результативності

LLM. Ці показники включають час обробки, оцінку релевантності рекомендацій та статистику використання API для кожної реалізації моделі.

Схема бази даних охоплює чотири основні сутності: Users, FairyTales, LLMModels та Recommendations. Зв'язки між цими сутностями структуровані таким чином, щоб підтримувати основну функціональність системи, полегшуючи при цьому аналіз ефективності.

Сутність FairyTales зберігає як завантажений користувачем контент, так і попередньо заповнену базу даних рекомендацій. Кожен запис містить поля метаданих, включаючи назву, автора, культурне походження, приблизний віковий діапазон, тематику та повний текст. Ці метадані підтримують як створення рекомендацій, так і подальшу оцінку якості рекомендацій.

Сутність LLMModels містить інформацію про конфігурацію для кожної інтегрованої мовної моделі, включно з кінцевими точками API, параметрами автентифікації та налаштуваннями обробки. Такий модульний підхід дозволяє додавати нові реалізації LLM без архітектурних модифікацій.

Сутність Recommendations записує всі згенеровані рекомендації, пов'язуючи вхідну казку, LLM, що рекомендує, рекомендовану казку, оцінку схожості та текст пояснення. Цей комплексний запис уможливорює поглиблений аналіз шаблонів рекомендацій і показників ефективності для різних реалізацій LLM.

Інтерфейс користувача використовує мінімалістичну філософію дизайну, що ставить на перше місце ефективність виконання завдань, водночас пристосовуючись до користувачів з різним рівнем технічної підготовки. Інтерфейс реалізує три основні робочі процеси взаємодії:

- завантаження контенту;
- вибір LLM;
- представлення результатів.

Інтерфейс представлення результатів використовує порівняльний перегляд, який підкреслює відмінності у виборі рекомендацій та поясненнях між LLM. Такий дизайн безпосередньо підтримує дослідницьку мету порівняльного аналізу ефективності LLM, роблячи розбіжності в рекомендаціях візуально очевидними.

Міркування доступності включають підтримку клавіатурної навігації, сумісність з пристроями для зчитування з екрану та принципи адаптивного дизайну, що враховують різноманітні потреби користувачів і контексти пристроїв. Ці міркування гарантують, що система залишається доступною для користувачів з різними здібностями та технічним середовищем.

Ця системна архітектура забезпечує структурну основу для впровадження системи рекомендацій казок, здатної підтримувати порівняльний аналіз різних підходів LLM, безпосередньо вирішуючи основні дослідницькі завдання щодо оцінки ефективності LLM у завданнях літературних рекомендацій.

## 2 ОГЛЯД ІСНУЮЧИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМ КАЗОК

Світ систем рекомендацій казок та дитячої літератури залишається надзвичайно малим, оскільки велика кількість існуючих систем зосереджуються на більш широкому освітньому контенті або загальних рекомендаціях щодо книг. Сучасні освітні системи рекомендацій в першу чергу орієнтовані на рекомендації щодо курсів, навчальні матеріали та загальні освітні ресурси, а не на спеціалізований літературний контент. Ця прогалина в літературі представляє як виклики, так і можливості для розробки цільових систем рекомендацій казок.

Найвідоміші сервіси з рекомендацій книг працюють на рівні загальної літератури, обслуговуючи різноманітну читацьку аудиторію без спеціалізації на дитячій літературі чи казках. Goodreads, що належить Amazon, є найбільшою соціальною платформою для читачів у світі, яка використовує підходи спільного фільтрування, що аналізують оцінки користувачів, історію читання та соціальні зв'язки для формування рекомендацій. Інтерфейс платформи зображений на рисунку 2.1.

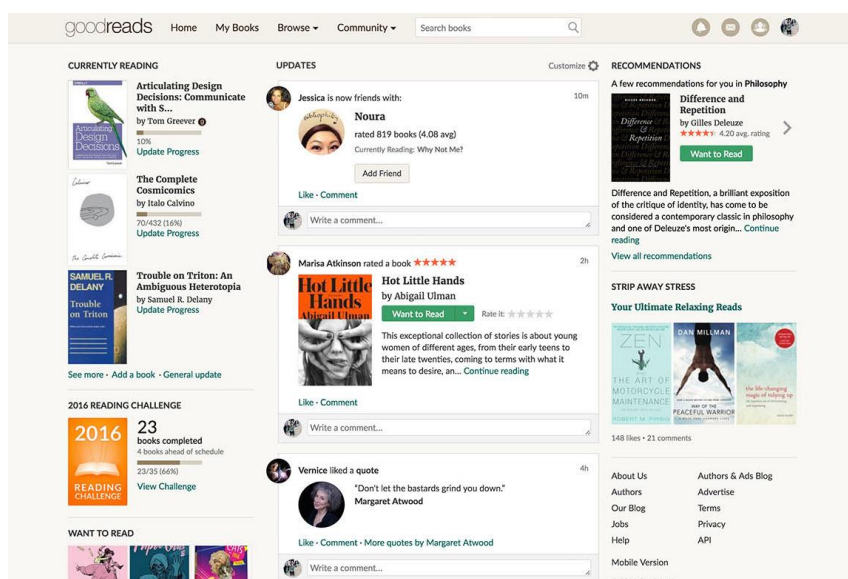


Рисунок 2.1 – Інтерфейс платформи Goodreads

Платформа обробляє дані мільйонів користувачів, щоб визначити читацькі вподобання та запропонувати книги на основі схожості з раніше оціненими творами. Однак Goodreads не має складних можливостей аналізу вмісту для специфічних особливостей казок, таких як структура оповіді, аналіз мотивів або фільтрування вмісту за віком.

Система рекомендацій книг Amazon використовує гібридний підхід, що поєднує спільну фільтрацію з методами на основі контенту, використовуючи історію покупок, поведінку користувачів під час перегляду та метадані товарів. Система Amazon використовує як моделі поведінки користувачів, так і метадані книг, включаючи жанр, автора та опис, для формування рекомендацій. Хоча ця система є ефективною для комерційних цілей, вона надає пріоритет показникам продажів над освітньою цінністю і не має розуміння наративу, яке є важливим для класифікації казок.

Наглядно гібридний підхід до рекомендації представлений на рисунку 2.2.

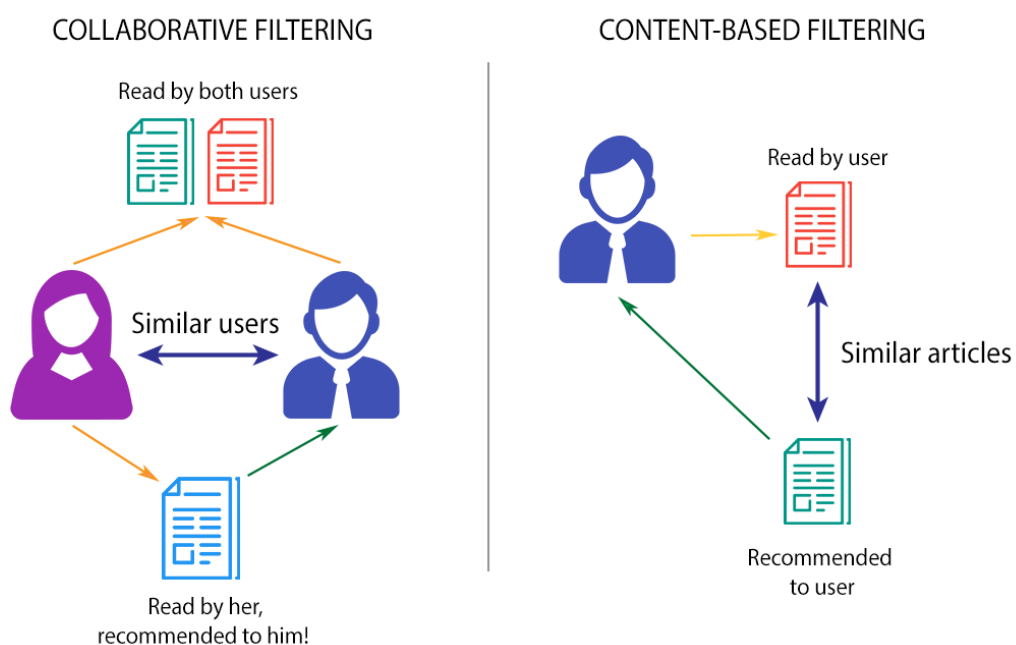


Рисунок 2.2 – Гібридний підхід до надання рекомендації

StoryGraph став помітною альтернативою, пропонуючи відстеження книг, функції спільноти та рекомендації з чіткими, візуально привабливими інтерфейсами, що дозволяють користувачам імпортувати дані Goodreads. Платформа забезпечує більш детальну категоризацію настрою та змісту, ніж традиційні системи, включаючи опції для читачів, щоб вказати бажаний емоційний тон та характеристики змісту, що продемонстровано на рисунку 2.3. Однак вона залишається орієнтованою переважно на літературу для дорослих та молоді, без спеціалізованих функцій для аналізу казок.

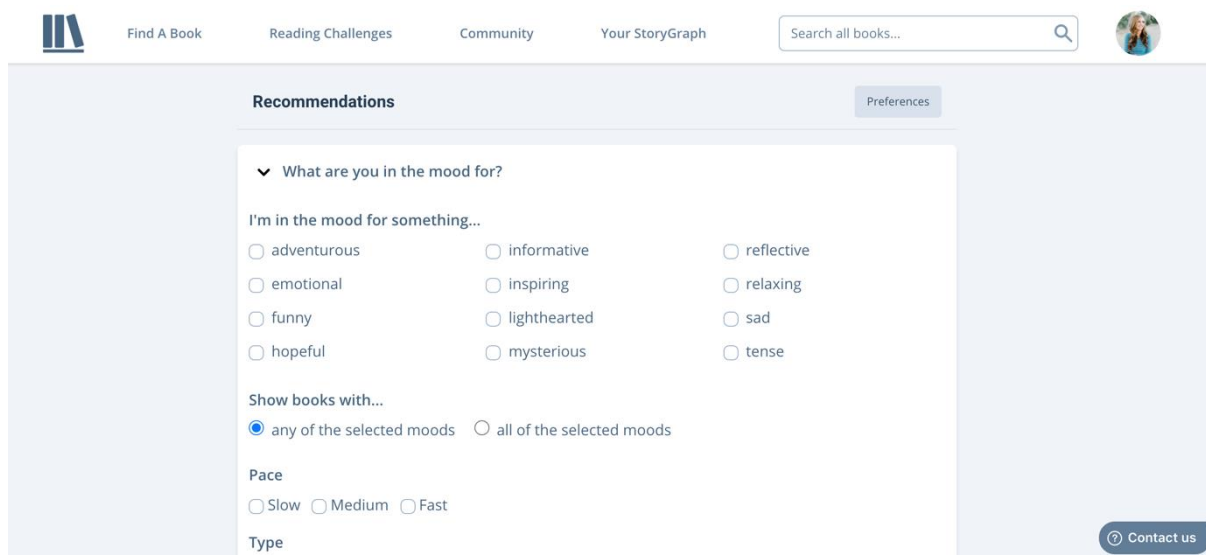


Рисунок 2.3 – Категоризація рекомендованої літератури за вподобаннями користувача на платформі StoryGraph

Системи рекомендацій в галузі освіти є більш релевантними для застосування в казках, хоча існуючі системи зосереджуються переважно на формальному освітньому контенті, а не на літературних матеріалах. Дослідження показують, що гібридні підходи стали провідною стратегією для створення рекомендацій в освітньому контексті, поєднуючи в собі кілька алгоритмічних технік для задоволення різноманітних навчальних потреб.

Системи управління навчанням (LMS), такі як Moodle, Canvas та Blackboard, містять базові функції рекомендацій, в першу чергу пропонуючи курси, завдання або навчальні ресурси на основі вимог навчальної програми та прогресу студентів. Ці системи мають на меті підтримати ініціативність учнів, полегшуючи доступ до відповідних ресурсів та зв'язків з однолітками. Однак їхні алгоритми рекомендацій розроблені для структурованого академічного контенту, а не для художньої літератури, що обмежує їх застосування в контексті казок.

Адаптивні платформи навчання, включаючи Khan Academy, Duolingo та подібні освітні технології, використовують складні алгоритми для персоналізації навчальних шляхів. Ці системи аналізують дані про успішність студентів, їхні навчальні уподобання та цілі навчальної програми, щоб рекомендувати відповідний контент. Хоча вони є технологічно досконалими, їм бракує семантичного розуміння, необхідного для аналізу літературного контенту, та культурної чутливості, необхідної для рекомендацій казок.

Декілька нішевих платформ намагалися вирішити проблему прогалин у літературних рекомендаціях, хоча жодна з них не зосереджується конкретно на казках або дитячій літературі.

What Should I Read Next? представляє підхід до рекомендацій на основі контенту, аналізуючи описи книг, жанри та теми, щоб запропонувати подібні назви. Платформа використовує алгоритми, які зосереджуються на настрої, характері, обстановці та характеристиках сюжету для генерації рекомендацій. Хоча цей підхід є більш досконалим, ніж базова спільна фільтрація, він залишається обмеженим у розумінні специфічних елементів казок, таких як фольклорні мотиви, культурні варіанти та відповідність віковим особливостям.

LibraryThing пропонує підхід до каталогізації та рекомендації книг, що базується на спільноті, з акцентом на бібліографічних даних та контенті, створеному користувачами. Платформа дозволяє детально тегувати книги

та пропонує алгоритми рекомендацій на основі бібліотечних колекцій, але не має аналітичної глибини, необхідної для специфічних особливостей казок.

Інтерфейс платформи разом із прикладом рекомендації книг на основі контенту у загальній секції Recommendations та на основі колаборативної фільтрації у секції Member Recommendations представлено на рисунку 2.4.

The screenshot displays the LibraryThing interface for the book "Harry Potter and the Sorcerer's Stone" by J.K. Rowling. The page is structured as follows:

- Header:** LibraryThing logo, navigation links (Home, Groups, Talk, More, Zeitgeist), and user options (Sign in / Join, English, Help).
- Book Title:** "Harry Potter and the Sorcerer's Stone" by J. K. Rowling.
- Series:** Harry Potter (1).
- Stats:** 144,001 Members, 2,125 Reviews, 4.29 average rating (4.29 stars), 101 Awards.
- On This Page:** A grid of links for Description, Members, Awards and Honors, Classifications, Tags, Lists, Series, and Statistics.
- Description:** A short synopsis: "Rescued from the outrageous neglect of his aunt and uncle, a young boy with a great destiny proves his worth while attending Hogwarts School for Wizards and Witches."
- Tags:** A collection of tags including "adventure", "British", "children", "children's", "children's books", "children's fiction", "children's literature", "England", "fantasy", "fantasy fiction", "fiction", "Harry Potter", "Harry Potter Series", "Hogwarts", "J.K. Rowling", "juvenile", "juvenile fiction", "magic", "movie", "Potter", "Rowling", "school", "series", "witchcraft", "witches", "wizard", "wizardry", "wizards", "YA", and "young adult".
- Recommendations:** A row of book covers with titles: "Harry Potter and the Chamber of Secrets", "Quidditch Through the Ages", "The Sorcerer's Companion: A Guide to the...", "Northern Lights", "Eragon", "The Subtle Knife", and "The Bad Beginning".
- Member Recommendations:** A grid of four book recommendations from members: "The Lightning Thief" by Rick Riordan, "The Lion, the Witch and the Wardrobe" by C. S. Lewis, "Artemis Fowl" by Eoin Colfer, and "So You Want to be a Wizard" by Diane Duane.
- Right Sidebar:**
  - Quick Facts:** Published 1997, Genres: Fantasy, Fiction and Literature, Kids, DDC/MDS: 823.92, LCC: PJ5124, Discussions: 1 current, 6 past.
  - Author:** J. K. Rowling, 305+ Works, 963,040 Members.
  - Popular Covers:** All 1,478 Covers >
  - Find It:** Amazon.com (direct), Library of Congress (direct).
  - Links:** Wikipedia.
  - Popularity:** A line graph showing popularity from 2011 to 2025.
  - Ratings:** Stars: 4.29 (4.29 stars), with a bar chart showing counts for each star rating (5: 17,106; 4: 10,015; 3: 4,219; 2: 863; 1: 263).
  - Helpers:** Helper Hub >

Рисунок 2.4 – Інтерфейс платформи LibraryThing разом із рекомендаціями до обраної книги

Whichbook пропонує інноваційний інтерфейс, що дозволяє користувачам вказувати бажані характеристики книги за допомогою повзунків, даючи читачам можливість вказати свої уподобання щодо емоційного тону, складності та тематичного змісту. Користувачі можуть вибирати характеристики персонажів, обставини історії та сюжетні моменти, щоб отримувати більш цілеспрямовані рекомендації, що продемонстровано на рисунку 2.5. Цей підхід демонструє потенціал для застосування до казок, але не має спеціалізованої бази знань, необхідної для точної категоризації казок.



Рисунок 2.5 – Надання рекомендованої літератури на основі вподобань користувача на платформі Whichbook

Академічні дослідження дали кілька експериментальних систем рекомендацій для літературного контенту, хоча лише деякі з них враховують специфічні вимоги до казок. Недавні дослідження, як наприклад у роботах [11], [12], [13], вивчали гібридні системи рекомендацій книг на основі шаблонів, використовуючи семантичні відносини, поєднуючи фільтрування на основі контенту з підходами спільного фільтрування. Ці системи демонструють підвищену точність порівняно з методами, що використовують один підхід, але залишаються зосередженими на загальній літературі, а не на спеціалізованих жанрах.

Дослідження фільтрування на основі контенту вивчали передові техніки обробки природної мови для рекомендацій книг, використовуючи такі методи, як векторизація TF-IDF, косинусна подібність та вбудовування слів для аналізу описів книг і генерації рекомендацій. Хоча ці підходи є технічно досконалыми, вони зазвичай зосереджуються на поверхневих текстових особливостях, а не на глибших нарративних структурах і культурних контекстах, які є важливими для аналізу казок.

Дослідження спільного фільтрування в освітньому контексті продемонстрували ефективність підходів на основі користувачів та елементів для рекомендацій навчальних матеріалів. Дослідження показують, що гібридні підходи, що поєднують спільні та контентні методи, досягають кращих результатів у сценаріях освітніх рекомендацій. Однак ці дослідження зазвичай зосереджуються на формальних освітніх матеріалах, а не на літературному контенті.

Аналіз існуючих систем виявляє кілька критичних обмежень у задоволенні вимог до рекомендацій казок:

- відсутність розуміння наративу;
- обмежена культурна чутливість;
- відсутність фільтрації за віком;
- відсутність інтеграції з освітою;
- недостатнє пояснення рекомендацій.

Сучасні системи не аналізують нарративні структури, фольклорні мотиви або архетипні шаблони, що характеризують казки. Існуючі системи рекомендацій книг зосереджуються переважно на поверхневих метаданих та моделях поведінки користувачів, а не на глибокому аналізі змісту.

Більшість платформ не мають механізмів для ідентифікації культурних варіантів подібних казок або забезпечення різноманітності в рекомендаціях. Це обмеження є особливо проблематичним для додатків з казками, де культурний контекст значно впливає на відповідність змісту та освітню цінність.

Хоча деякі системи пропонують базову категоризацію за жанрами, жодна з них не забезпечує складного аналізу відповідності розвитку, оцінки рівня читання або оцінки чутливості змісту, що є надзвичайно важливим для дитячої літератури.

Існуючі системи не включають педагогічні рамки, узгодження з навчальною програмою або підтримку освітніх цілей, що підвищило б їхню корисність в освітньому контексті.

Більшість платформ надають мінімальне пояснення обґрунтування рекомендацій, не маючи детального обґрунтування, необхідного для освітніх додатків, де прозорість і підзвітність є надзвичайно важливими.

### 3 АНАЛІТИЧНИЙ АНАЛІЗ LLM ТА РОЗРОБКА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ КАЗОК

Впровадження порівняльного аналітичного аналізу LLM передбачає використання модульної, об'єктно-орієнтованої архітектури, розробленої для полегшення систематичної оцінки декількох мовних моделей для вбудовування текстів казок. Філософія проектування системи наголошує на гнучкості, розширюваності та всебічних можливостях оцінки, зберігаючи при цьому чітке розділення функціональних областей.

Основна реалізація складається з чотирьох основних компонентів, які працюють разом, щоб забезпечити повну систему оцінки: `BookEmbeddingSystem` для генерації вбудовування та аналізу схожості, `EmbeddingEvaluator` для кількісної оцінки продуктивності, `EvaluationVisualizer` для неконтрольованого представлення результатів та `SupervisedEvaluationVisualizer` для аналізу на основі фактичних даних. Цей модульний підхід дозволяє проводити незалежне тестування та валідацію кожного компонента, підтримуючи при цьому безперервну інтеграцію для всебічного порівняння моделей.

Система підтримує як локальне розгортання через Ollama, так і хмарні моделі через трансформери HuggingFace, забезпечуючи гнучкість у виборі моделі та використанні обчислювальних ресурсів. Цей підхід з подвійним бекендом враховує практичні обмеження дослідницьких середовищ, де доступ до обчислювальних ресурсів може змінюватися, забезпечуючи при цьому відтворюваність у різних сценаріях розгортання.

Клас `BookEmbeddingSystem` служить центральним координатором для генерації вбудовування, реалізуючи гнучку архітектуру бекенду, яка безперешкодно інтегрує декілька постачальників моделей. Метод ініціалізації демонструє цей підхід до проектування, що наведений у лістингу 3.1.

### Лістинг 3.1 – Програмний код ініціалізації екземплярів класу `BookEmbeddingSystem`

```

def __init__(self, dataset_file: str, model: str = "all-
minilm",
             books_directory: str = "books", ollama_url:
str = "http://localhost:11434",
             model_type: str = "ollama"):
    self.model_type = model_type.lower()
    self.hf_model = None
    self.hf_tokenizer = None

    if self.model_type == "huggingface":
        self._initialize_huggingface_model()
    elif self.model_type == "ollama":
        logger.info(f"Initialize System with Ollama model
{model}")
    else:
        raise ValueError(f"Unsupported model_type:
{model_type}")

```

Ця стратегія впровадження враховує неоднорідний характер сучасного розгортання LLM, де дослідникам може знадобитися оцінка моделей на різних платформах і в різних обчислювальних середовищах. Система автоматично обробляє складнощі ініціалізації, характерні для кожного бекенду, одночасно надаючи уніфікований інтерфейс для генерації вбудовування.

Процес генерації вбудовування включає в себе складну обробку помилок і попередню обробку тексту, щоб забезпечити надійну роботу з різними текстами казок, що мають різні формати та кодування. Метод `read_book_content` є прикладом такого підходу, що наведений у лістингу 3.2.

### Лістинг 3.2 – Програмний код методу зчитування вмісту казки

```

def read_book_content(self, file_path: Path) -> str:

```

### Продовження лістингу 3.2

```

    try:
        with open(file_path, 'r', encoding='utf-8') as
file:
            content = file.read()
            return content
    except UnicodeDecodeError:
        try:
            with open(file_path, 'r', encoding='latin-1')
as file:
                content = file.read()
                return content
        except Exception as e:
            logger.error(f"Error reading {file_path}:
{e}")
            return ""

```

Ця стратегія резервного кодування враховує реалії роботи з цифровими колекціями казок, які можуть походити з різних джерел з неоднаковими стандартами кодування символів. Реалізація надає пріоритет збереженню даних, одночасно підтримуючи стабільність системи за допомогою коректної обробки помилок.

Сам процес генерації вбудовування демонструє ретельну увагу до обмежень, характерних для конкретної моделі, та стратегій оптимізації. Для моделей HuggingFace реалізація включає в себе складну токенізацію та маскування уваги, яке наведено у лістингу 3.3.

### Лістинг 3.3 – Програмний код ініціалізації моделей HuggingFace

```

def _get_embedding_from_huggingface(self, text: str) ->
List[float]:
    if hasattr(self.hf_model, 'encode'):
        # Using sentence-transformers
        max_length = 512

```

### Продовження лістингу 3.3

```

        if len(text) > max_length * 4:
            text = text[:max_length * 4]

        embedding = self.hf_model.encode(text,
convert_to_tensor=False)
            return embedding.tolist() if hasattr(embedding,
'tolist') else list(embedding)
        else:
            # Using transformers library with attention
masking
            inputs = self.hf_tokenizer(text,
return_tensors='pt', truncation=True,
padding=True,
max_length=max_length)

            with torch.no_grad():
                outputs = self.hf_model(**inputs)
                # Mean pooling with attention mask
                masked_embeddings = outputs.last_hidden_state
* inputs['attention_mask'].unsqueeze(-1)
                mean_embedding = torch.sum(masked_embeddings,
dim=1) / torch.sum(inputs['attention_mask'], dim=1,
keepdim=True)

                return mean_embedding.squeeze().tolist()

```

Така реалізація демонструє глибоке розуміння нюансів архітектури трансформерів, реалізуючи належну маскуванню уваги та стратегії об'єднання середніх значень, що максимізують якість витягнутих вбудовувань для подальшого аналізу схожості.

Система використовує кілька показників схожості, щоб забезпечити комплексні можливості порівняння, враховуючи, що різні показники схожості можуть бути оптимальними для різних аспектів аналізу казок.

Модульна конструкція дозволяє легко розширювати систему додатковими показниками, яка наведена у лістингу 3.4.

#### Лістинг 3.4 – Програмний код аналізу схожості казок

```
def find_similar_books(self, query_embedding: List[float],
                      similarity_method: str = "cosine",
                      top_k: int = 10) -> List[Tuple[str,
float]]:
    if similarity_method == "cosine":
        return
self.cosine_similarity_search(query_embedding, top_k)
    elif similarity_method == "euclidean":
        return
self.euclidean_similarity_search(query_embedding, top_k)
    elif similarity_method == "dot_product":
        return
self.dot_product_similarity_search(query_embedding, top_k)
```

Кожна реалізація методу подібності включає відповідну нормалізацію та масштабування, щоб забезпечити значуще порівняння між різними просторами вбудовування та архітектурами моделей. Наприклад, реалізація косинусної подібності використовує оптимізовані алгоритми scikit-learn, зберігаючи сумісність із ширшою системою оцінки.

Клас `EmbeddingEvaluator` реалізує комплексний набір метрик безконтрольної (unsupervised) оцінки, які оцінюють якість вбудовування без необхідності ручного анотування даних базової істини. Цей підхід є особливо цінним для аналізу казок, де отримання експертних анотацій для великих колекцій може бути надто дорогим і суб'єктивним.

Набір безконтрольної оцінки включає оцінку якості кластеризації, аналіз найближчих сусідів, виявлення винятків та аналіз простору вбудовування. Загальний процес безконтрольної оцінки наведений у лістингу 3.5.

## Лістинг 3.5 – Програмний код проведення безконтрольної оцінки

```
def unsupervised_evaluation_suite(self) -> Dict[str, Any]:
    results = {}
    print("Running unsupervised evaluation suite...")
    # Embedding space analysis
    results['embedding_space'] =
self.analyze_embedding_space()
    # Nearest neighbor analysis
    results['nearest_neighbors'] =
self.analyze_nearest_neighbors()
    # Unsupervised clustering
    results['clustering'] =
self.evaluate_unsupervised_clustering()
    # Outlier detection
    results['outliers'] = self.detect_outliers()

    return results
```

Цей комплексний підхід дозволяє оцінити якість вбудовування за різними параметрами, надаючи інформацію як про локальну структуру, так і про глобальну організацію простору вбудовування.

Компонент оцінки кластеризації реалізує декілька алгоритмів і метрик для оцінки того, наскільки добре вбудовування відображає значущі групування в колекціях казок. Реалізація демонструє складну методологію вибору параметрів та оцінки. Результати аналізу зберігаються в об'єкті, структура якого наведена у лістингу 3.6.

Лістинг 3.6 – Програмний код ініціалізації об'єкту результатів кластерного аналізу

```
results = { 'silhouette_scores': {},
'calinski_harabasz_scores': {}, 'davies_bouldin_scores': {},
'inertias': {}, 'best_n_clusters': {} }
```

Цей мультиметричний підхід забезпечує надійну оцінку якості кластеризації, причому кожна метрика фіксує різні аспекти валідності кластерів. Реалізація автоматично визначає оптимальну кількість кластерів за різними метриками, що дозволяє провести комплексне порівняння того, як різні LLM організують зміст казок.

Система візуалізації реалізує складну структуру презентації, яка перетворює складні результати числової оцінки в зрозумілі візуальні зображення. Клас `EvaluationVisualizer` демонструє розширену інтеграцію `matplotlib` і `seaborn` для створення зображень високої якості.

Дизайн інформаційної панелі відповідає принципам візуалізації інформації, представляючи ієрархічну інформацію від загальних підсумків до детальних розбивок показників. Реалізація включає інтерактивні елементи та можливості експорту, що підтримують як дослідницький аналіз, так і вимоги до офіційної звітності.

Фреймворк візуалізації включає спеціалізовану функціональність для порівняння продуктивності декількох LLM за стандартизованими показниками. Реалізація порівняльного аналізу демонструє складні стратегії агрегації та представлення даних, що дозволяє дослідникам визначити найефективніший LLM для конкретних аспектів аналізу казок, зберігаючи при цьому комплексний контроль за ефективністю моделі за всіма параметрами оцінки.

Впровадження включає надійні стратегії управління даними, що підтримують відтворювані дослідницькі практики. Вбудована система кешування демонструє продуманий баланс між обчислювальною ефективністю та експериментальною гнучкістю, що продемонстроване у лістингу 3.7.

Лістинг 3.7 – Програмний код завантаження датасету

```
def load_dataset(self, dataset_dir: str = "embeddings") ->
bool:
```

### Продовження лістингу 3.7

```
path = os.path.join(dataset_dir, self.dataset_file)
try:
    if os.path.exists(path):
        with open(path, 'rb') as f:
            self.embeddings_data = pickle.load(f)
            logger.info(f"Dataset loaded from {path}")
            return True
    else:
        logger.warning(f"Dataset file {path} not
found")
        return False
```

Цей механізм кешування значно зменшує обчислювальні накладні витрати для ітеративних експериментів, зберігаючи цілісність даних і забезпечуючи відтворюваність результатів у різних експериментальних сесіях.

Реалізація демонструє витончені практики програмної інженерії, що підходять для дослідницьких контекстів, включаючи комплексну обробку помилок, системи реєстрації та модульну конструкцію, що полегшує як тестування окремих компонентів, так і оцінку інтегрованої системи. Гнучкість фреймворку підтримує різноманітні дослідницькі питання, зберігаючи при цьому необхідну строгість для систематичного порівняльного аналізу продуктивності LLM у спеціалізованих літературних галузях.

Для проведення аналітичного аналізу LLM моделей було взято три наступні:

- all-MiniLM-L6-v2;
- Nomic Embed Text v1;
- DeepSeek-R1-Distill-Qwen-1.5B.

All-MiniLM-L6-v2 – це модель трансформаторів речень, розроблена Microsoft і спільнотою Hugging Face, яка відображає речення та абзаци у

384-вимірному щільному векторному просторі. Ця модель була налагоджена на основі попередньо навченої моделі preimers/MiniLM-L6-N384-uncased на наборі даних із 1 мільярдом пар речень з використанням контрастної навчальної мети. Модель використовує архітектуру на основі BERT, оптимізовану для ефективності, що робить її придатною для завдань семантичної схожості, кластеризації та пошуку інформації.

У процесі навчання використовувалася контрастна навчальна структура, в якій модель вчиться передбачати, які речення були фактично згруповані в пари з набору випадково відібраних альтернатив. Такий підхід дозволяє моделі фіксувати багаті семантичні відносини, зберігаючи при цьому обчислювальну ефективність. Компактний розмір і надійна продуктивність моделі роблять її популярним вибором для додатків, що вимагають швидкого виведення з розумним семантичним розумінням.

Nomic Embed Text v1 є значним проривом у галузі моделей вбудовування тексту з відкритим кодом, що має довжину послідовності 8192, яка перевершує text-embedding-ada-002 та text-embedding-v3-small від OpenAI як у коротких, так і в довгих контекстних завданнях. Розроблена Nomic AI, ця модель усуває обмеження існуючих вбудованих моделей з відкритим кодом, забезпечуючи конкурентоспроможну продуктивність з власними альтернативами, зберігаючи при цьому повну прозорість і відтворюваність.

Модель використовує багатоступеневий конвеєр навчання, починаючи з неконтрольованого контрастного навчання на слабо пов'язаних парах текстів (таких як пари питання-відповідь з форумів, пари заголовків-текстів з оглядів і резюме новин), а потім тонке налаштування на позначених наборах даних вищої якості. Унікальною особливістю цієї моделі є її вимога щодо префіксів, специфічних для завдань (таких як «search\_document:», «search\_query:», «classification:», «clustering:»), які інструктують модель про передбачуваний випадок використання, оптимізуючи продуктивність для конкретних застосувань.

DeepSeek-R1-Distill-Qwen-1.5B – це дистильована версія більшої моделі міркування DeepSeek-R1, створена шляхом дистиляції знань з використанням приблизно 800 000 високоякісних зразків, згенерованих повномасштабною моделлю DeepSeek-R1. Ця модель з 1,5 мільярдами параметрів базується на архітектурі Qwen-2.5 і включає в себе розширені можливості міркування завдяки обробці ланцюжка думок.

На відміну від традиційних моделей вбудовування, DeepSeek-R1 була розроблена в першу чергу для завдань міркування і використовує методи навчання на основі підкріплення. Модель демонструє складні здібності до вирішення проблем за допомогою видимого міркування ланцюжком думок, самоперевірки та багатоетапних можливостей рефлексії. Важливо зазначити, що хоча ця модель може генерувати вбудовування, вона була спеціально оптимізована для завдань міркування, а не для застосувань семантичної схожості, що може вплинути на її продуктивність у традиційних тестах оцінки вбудовування.

Основні базові параметри моделей наведені у таблиці 3.1.

Таблиця 3.1 – Ключові метрики моделей для створення текстових вбудувань

Model	Parameters	Architecture	Embedding Dimension	Context Length	Training Method
all-MiniLM-L6-v2	~22.7M	BERT-based (MiniLM-L6)	384	256 word pieces	Contrastive learning on 1B pairs
Nomic Embed Text v1	~137M	Modified BERT-base	768 (Matryoshka)	8,192 tokens	Multi-stage contrastive + supervised
DeepSeek-R1-Distill-1.5B	1.5B	Qwen-2.5 based	Variable*	32,768 tokens	Knowledge distillation + RL

Аналіз українських казок за допомогою системи класифікації ATU (Aarne-Thompson-Uther) є чудовою основою для контрольованої (supervised) оцінки ефективності моделі. Ієрархічна структура типів казок індексу ATU пропонує кілька рівнів семантичної

деталізації, що дозволяє ефективно перевірити, наскільки добре кожна модель вбудовування відображає як поверхневі елементи розповіді, так і глибші структурні закономірності змісту казок.

Ця порівняльна основа дозволяє оцінити, чи можуть компактні моделі, такі як all-MiniLM-L6-v2, конкурувати з більшими, складними архітектурами для літературного аналізу в конкретній галузі, а також дослідити, як моделі, оптимізовані для міркування, такі як DeepSeek-R1, працюють, коли їх адаптують для завдань вбудовування поза межами їх основного призначення.

Оцінка показує принципово різні підходи до вбудовування організації простору в трьох моделях. Моделі all-MiniLM-L6-v2 і Nomic Embed Text демонструють надзвичайно схожі властивості вбудовування простору, з середньою схожістю 0,783 і 0,782 відповідно. Така висока схожість свідчить про те, що ці моделі створюють відносно щільні простори вбудовування, де казки групуються близько одна до одної, що потенційно вказує на те, що вони фіксують широкі семантичні схожості, але можуть мати труднощі з розрізненням нюансів між казками.

Результати аналізу моделі all-MiniLM-L6-v2 продемонстровані на рисунку 3.1.

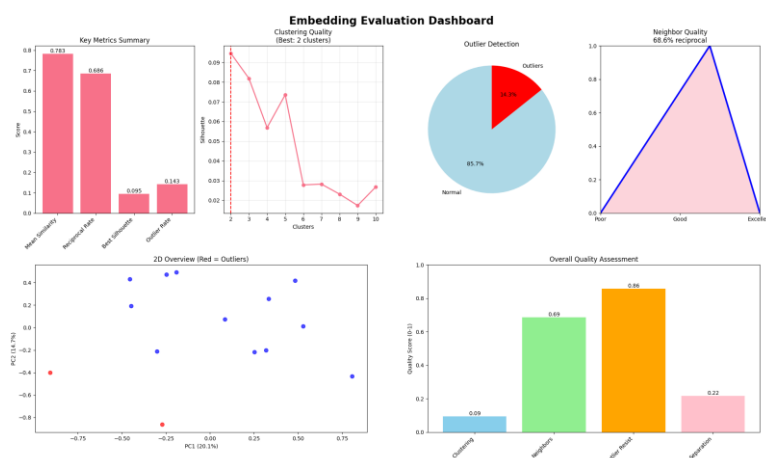


Рисунок 3.1 – Результати аналізу моделі all-MiniLM-L6-v2

Результати аналізу моделі Nomic Embed Text продемонстровані на рисунку 3.2.

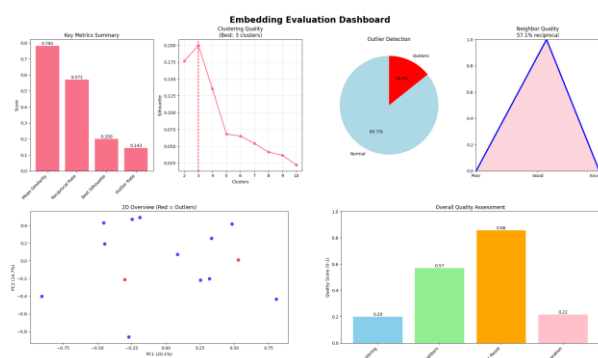


Рисунок 3.2 – Результати аналізу моделі Nomic Embed Text

На противагу цьому, DeepSeek-R1 створює значно більш розсіяний простір вбудовування із середньою схожістю лише 0,508 і набагато вищим стандартним відхиленням (0,283 проти 0,048–0,067 для інших). Ця нижча базова схожість у поєднанні з вищою мінливістю свідчить про те, що модель створює більш диференційовані представлення, які краще зберігають унікальні характеристики окремих казок, а не групують їх у надто схожі кластери.

Результати аналізу моделі DeepSeek-R1 продемонстровані на рисунку 3.3.

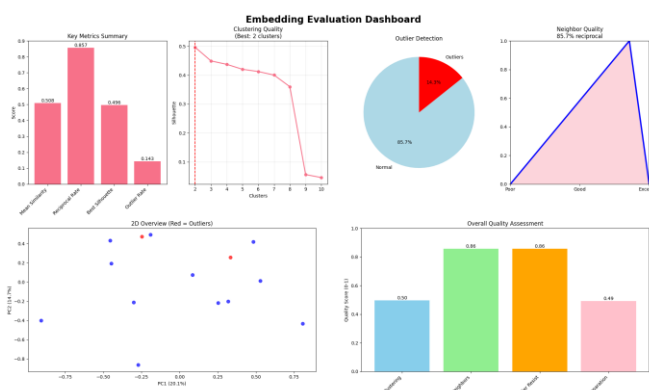


Рисунок 3.3 – Результати аналізу моделі DeepSeek-R1

Оцінка кластеризації виявляє найістотніші відмінності в продуктивності між моделями. DeepSeek-R1 досягає показника силуету 0,496, що є суттєвим поліпшенням порівняно з показником all-MiniLM-L6-v2 (0,095) та Nomic Embed (0,200). Хоча жодна з моделей не досягає справді відмінної продуктивності кластеризації (зазвичай вище 0,7), показник DeepSeek вказує на значуще розділення кластерів, яке може ефективно підтримувати завдання класифікації ATU.

Вища ефективність кластеризації, ймовірно, походить від методології навчання DeepSeek, орієнтованої на міркування. На відміну від традиційних моделей вбудовування, оптимізованих для загальної семантичної схожості, підхід DeepSeek до підкріплювального навчання, здається, розвиває представлення, які краще відображають структурні та тематичні відмінності, що мають вирішальне значення для розрізнення типів казок. Це відкриття свідчить про те, що моделі, навчені міркуванням, можуть надавати несподівані переваги для завдань літературного аналізу, що вимагають тонких категоріальних розрізень.

Аналіз взаємних сусідів, мабуть, дає найбільш показове уявлення про поведінку моделі. DeepSeek-R1 демонструє виняткову узгодженість сусідів із взаємним коефіцієнтом 85,7%, що означає, що коли казка А вважає казку Б одним із своїх найближчих сусідів, Б у відповідь на це у більшості випадків також вважає казку А своїм найближчим сусідом. Ця двостороння узгодженість вказує на стабільні, значущі відносини подібності в просторі вбудовування.

І навпаки, all-MiniLM-L6-v2 досягає помірною коефіцієнта взаємності 68,6%, тоді як Nomic Embed Text демонструє тривожну асиметрію на рівні лише 57,1%. Ця асиметрія в Nomic Embed свідчить про те, що модель може давати суперечливі оцінки схожості, де відносини між казками значно різняться залежно від напрямку порівняння, що може становити потенційну проблему надійності для систематичного аналізу.

Аналіз розмірності дав цікавий результат. Незважаючи на найвищу розмірність вбудованого простору (1536), DeepSeek-R1 потребує найменшої кількості розмірностей для охоплення 90% і 95% дисперсії (відповідно 6 і 7 розмірностей). Така ефективність свідчить про те, що модель ефективніше використовує свій високорозмірний простір, зосереджуючи значущу інформацію в меншій кількості ключових розмірностей, а решту простору використовуючи для дрібних відмінностей.

Як all-MiniLM-L6-v2, так і Nomic Embed Text потребують 10–12 вимірів для фіксації еквівалентної дисперсії, що вказує на те, що їх розподіл інформації є більш розсіяним у просторі вбудовування. Ця різниця може відображати структурований підхід DeepSeek до навчання представлення через його орієнтований на міркування процес навчання.

Цікаво, що всі три моделі визначають однакові показники відхилень (14,3%), хоча вони позначають різні конкретні казки як найбільш екстремальні відхилення. Така узгодженість свідчить про справжні структурні відмінності в збірці казок, які виходять за межі архітектури окремих моделей. Визначення відхилень дає цінну інформацію про казки, які можуть представляти унікальні нарративні структури або тематичні елементи в українській збірці.

На основі комплексних показників оцінки DeepSeek-R1-Distill-Qwen-1.5B виявляється найкращою моделлю для аналізу вбудовування українських казок. Незважаючи на те, що вона розроблена в першу чергу для міркувань, а не для завдань вбудовування, вона демонструє:

- вищу якість кластеризації з показником силуету, майже в 2,5 рази вищим за найкращу альтернативу;
- виняткову узгодженість сусідів на рівні 85,7%, що свідчить про надійні відносини схожості;
- більш різноманітний та інформативний простір вбудовування з нижчою середньою схожістю, але вищою дискримінаційною здатністю;

– більша ефективність вимірювання, незважаючи на більший власний простір вбудовування.

Загальне порівняння метрик між трьома моделями наведено на рисунку 3.4.

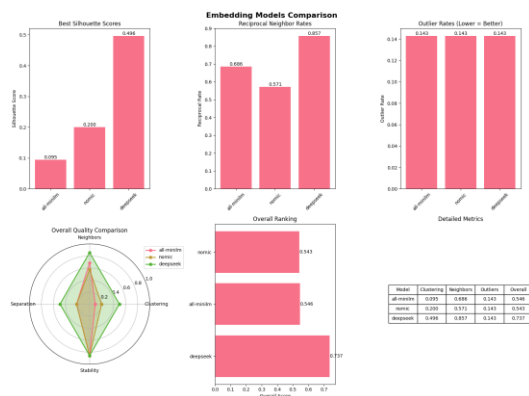


Рисунок 3.4 – Загальні аналітичного аналізу трьох моделей

Результати свідчать про те, що методи навчання, орієнтовані на міркування, можуть надати несподівані переваги для застосувань літературного аналізу. Здатність DeepSeek розробляти структуровані, ієрархічні представлення за допомогою підкріплювального навчання видається особливо придатною для тонких категоріальних відмінностей, необхідних для систем класифікації казок, таких як індекс ATU.

Хоча all-MiniLM-L6-v2 та Nomic Embed Text демонструють адекватну ефективність для загальних завдань семантичної схожості, їхня схильність до високої базової схожості може обмежувати їхню ефективність для детального літературного аналізу, де збереження унікальних нарративних характеристик є надзвичайно важливим. Вища ефективність DeepSeek-R1 підтверджує гіпотезу, що вдосконалені можливості міркування можуть перетворитися на більш складне розуміння тексту, навіть у спеціалізованих додатках, що виходять за межі основного призначення моделі.

## 4 РОЗРОБЛЕНА РЕКОМЕНДАЦІЙНА СИСТЕМА КАЗОК

Система рекомендацій казок пропонує користувачам інтуїтивно зрозумілий веб-інтерфейс, розроблений за принципом поступового розкриття інформації, де складність вводиться поступово, у міру того як користувачі все глибше занурюються в систему. Додаток використовує послідовну парадигму навігації завдяки адаптивному дизайну на основі Bootstrap, який безперешкодно адаптується до настільних і мобільних пристроїв.

Основна панель навігації, що має градієнтну колірну гаму від фіолетового до синього, забезпечує постійний доступ до чотирьох основних функціональних областей: Головна, Пошук подібних, Бібліотека та Додати книгу. Ця структура навігації відображає типовий шлях користувача від відкриття до активної взаємодії з системою рекомендацій, підтримуючи як випадкове дослідження, так і систематичне створення бібліотеки.

Початкова сторінка слугує як вступ, так і центром заклику до дії, використовуючи візуальну ієрархію для привернення уваги користувачів, інтерфейс якої зображений на рисунку 4.1.

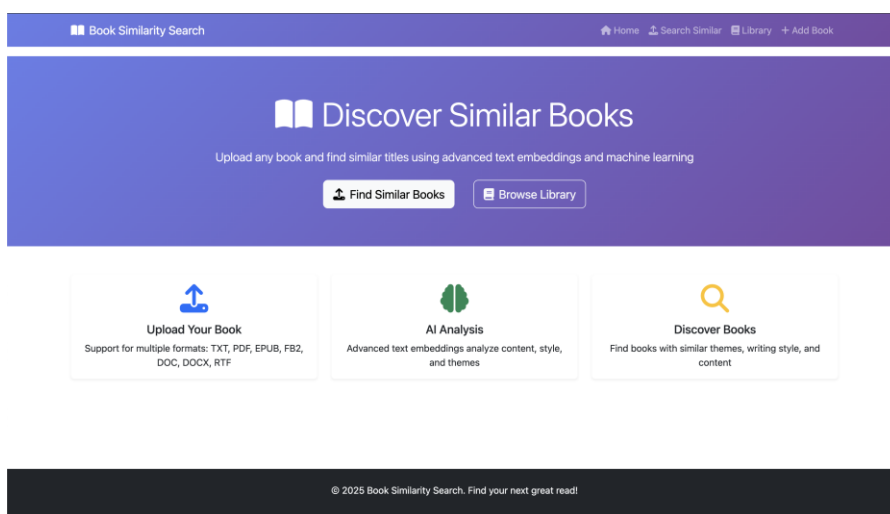


Рисунок 4.1 – Інтерфейс головної сторінки

Виділена секція містить головну цінність пропозиції – «Знайти подібні книги» – супроводжувану пояснювальним текстом про можливості машинного навчання системи. Інтерфейс навмисно уникає технічного жаргону, натомість зосереджується на перевагах та результатах для користувачів.

Три картки функцій під розділом використовують іконографічне представлення для пояснення процесу рекомендацій: завантаження, аналіз AI та відкриття. Ця послідовність візуальних метафор допомагає користувачам зрозуміти робочий процес системи без необхідності детальних технічних знань. Дві основні кнопки дій – «Знайти подібні книги» та «Переглянути бібліотеку» – забезпечують чіткі точки входу для різних намірів користувачів.

Сторінка завантаження представляє основну цінність системи за допомогою складного інтерфейсу перетягування, який підтримує різні формати файлів.

Область завантаження має механізми візуального зворотного зв'язку, включаючи стан наведення курсору та індикатори перетягування, які забезпечують миттєву реакцію на дії користувача. Підтримка різних форматів книг (TXT, PDF, EPUB, FB2, DOC, DOCX, RTF) демонструє практичну користь системи для користувачів з різноманітними цифровими бібліотеками.

Інтерфейс включає параметри рекомендацій за допомогою двох випадючих меню: метод схожості (косинусна схожість, евклідова відстань або скалярний добуток) та кількість результатів (5–20 книг). Хоча ці опції надають досвідченим користувачам можливість точного контролю, стандартні налаштування забезпечують можливість роботи з системою навіть для початківців, які не розуміють математичних концепцій, що лежать в її основі. Загальний вигляд форми представлений на рисунку 4.2.

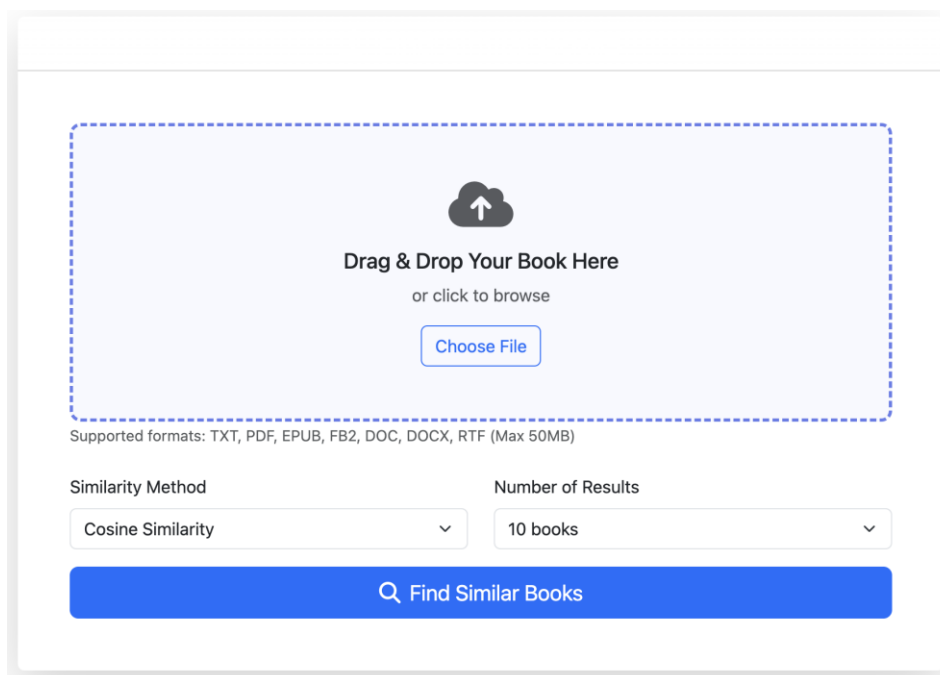


Рисунок 4.2 – Інтерфейс форми для пошуку схожих казок

Візуальний зворотний зв'язок під час обробки досягається за допомогою динамічних станів кнопок та індикаторів завантаження. Кнопка відправлення перетворюється з «Знайти подібні книги» на «Обробка...» з анімованим обертвовим індикатором, що чітко вказує на те, що система активно працює. Такий підхід зменшує занепокоєння користувачів під час потенційно тривалих процесів генерації вбудовування та обчислення схожості.

Сторінка бібліотеки функціонує як браузер сховища та як відправна точка для порівняльного аналізу.

Книги відображаються в адаптивному макеті сітки за допомогою компонентів на основі карток, які містять основні метадані: назву, автора, зображення обкладинки (якщо є) та обсяг вмісту. Ефекти наведення курсору на картку створюють інтерактивне відчуття, а послідовна візуальна обробка встановлює шаблони, які користувачі швидко освоюють для навігації.

Інтерфейс сторінки із казками представлено на рисунку 4.3.

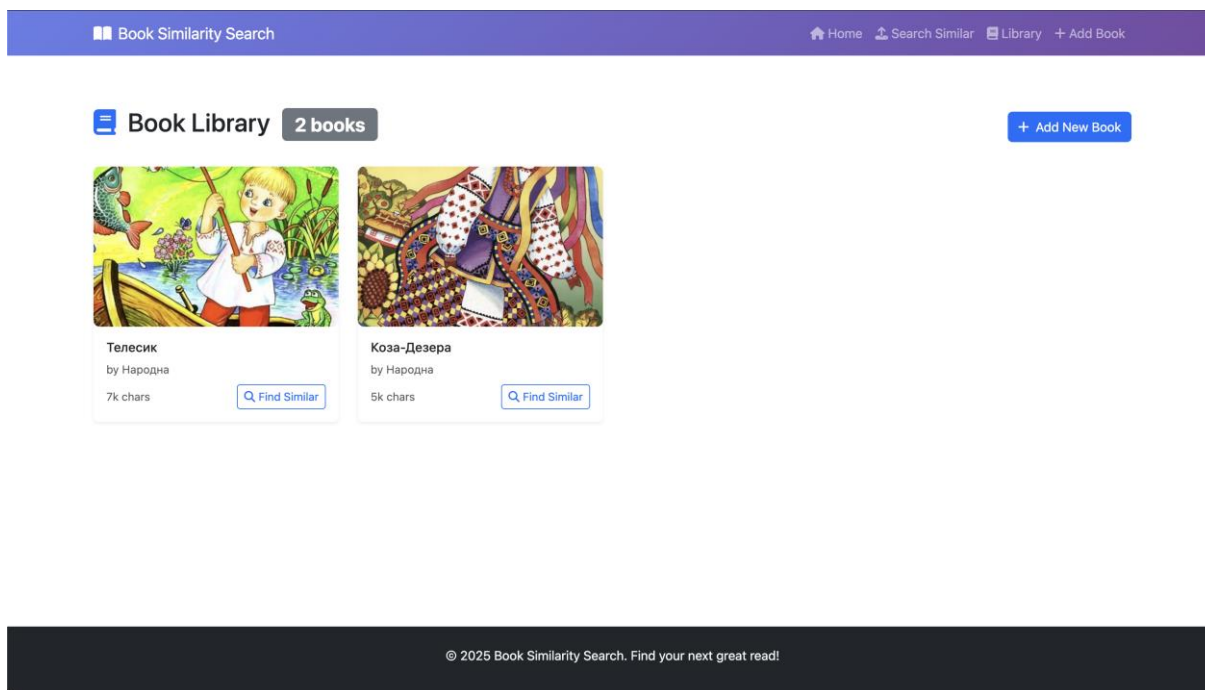


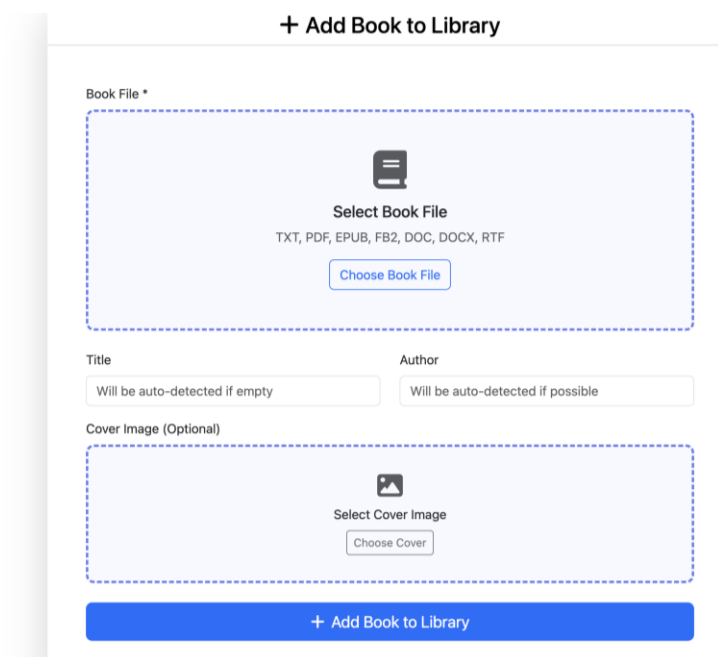
Рисунок 4.3 – Інтерфейс бібліотеки

Кожна картка книги містить кнопку дії «Знайти подібне», що дозволяє користувачам знаходити рекомендації на основі книг, які вже є в їхній колекції. Ця функція перетворює бібліотеку зі статичного сховища на активний інструмент пошуку, заохочуючи до дослідження та випадкових зустрічей із пов'язаним контентом.

Для колекцій без зображень обкладинки система використовує естетично узгоджену графіку-заповнювач із піктограмами книг на градієнтному тлі. Такий дизайн забезпечує візуальну узгодженість, уникаючи при цьому ефекту розриву через відсутність зображень або загальних заповнювачів.

Функція додавання книг демонструє вишуканий дизайн користувацького досвіду завдяки двофункціональному інтерфейсу, який підтримує як завантаження контенту, так і збагачення метаданих. Сторінка використовує дві окремі області завантаження – одну для контенту книги, а іншу для зображень обкладинок – кожна з яких має індивідуальне візуальне відображення та перевірку типу файлу.

Інтерфейс форми для завантаження казок представлений на рисунку 4.4.



The screenshot shows a web form titled '+ Add Book to Library'. It contains the following elements:

- Book File \***: A dashed blue box containing a document icon, the text 'Select Book File', supported file formats (TXT, PDF, EPUB, FB2, DOC, DOCX, RTF), and a 'Choose Book File' button.
- Title**: A text input field with the placeholder text 'Will be auto-detected if empty'.
- Author**: A text input field with the placeholder text 'Will be auto-detected if possible'.
- Cover Image (Optional)**: A dashed blue box containing an image icon, the text 'Select Cover Image', and a 'Choose Cover' button.
- Submit Button**: A large blue button at the bottom with the text '+ Add Book to Library'.

Рисунок 4.4 – Інтерфейс форми для завантаження казок

Поля введення метаданих для назви та автора містять інтелектуальний текст-заповнювач, який вказує, що інформація буде «автоматично виявлена, якщо поле порожнє», формуючи відповідні очікування користувачів щодо можливостей системи автоматичного аналізу вмісту. Такий підхід зменшує навантаження на користувача, зберігаючи можливість ручного втручання, якщо автоматичне виявлення виявляється недостатнім.

Візуальна зворотний зв'язок під час процесу завантаження включає відображення розміру файлу, індикатори прогресу та динамічних станів кнопок та індикаторів завантаження. Кнопка відправлення перетворюється із «Знайти подібні книги» на «Обробка...» з анімованим обертовим індикатором, що чітко вказує на те, що система активно працює. Такий підхід зменшує занепокоєння користувачів під час потенційно тривалих процесів генерації вбудовування та обчислення схожості.

Сторінка результатів є кульмінацією процесу рекомендацій завдяки візуально насиченому представленню подібних книг. Інтерфейс цієї сторінки наведено на рисунку 4.5.

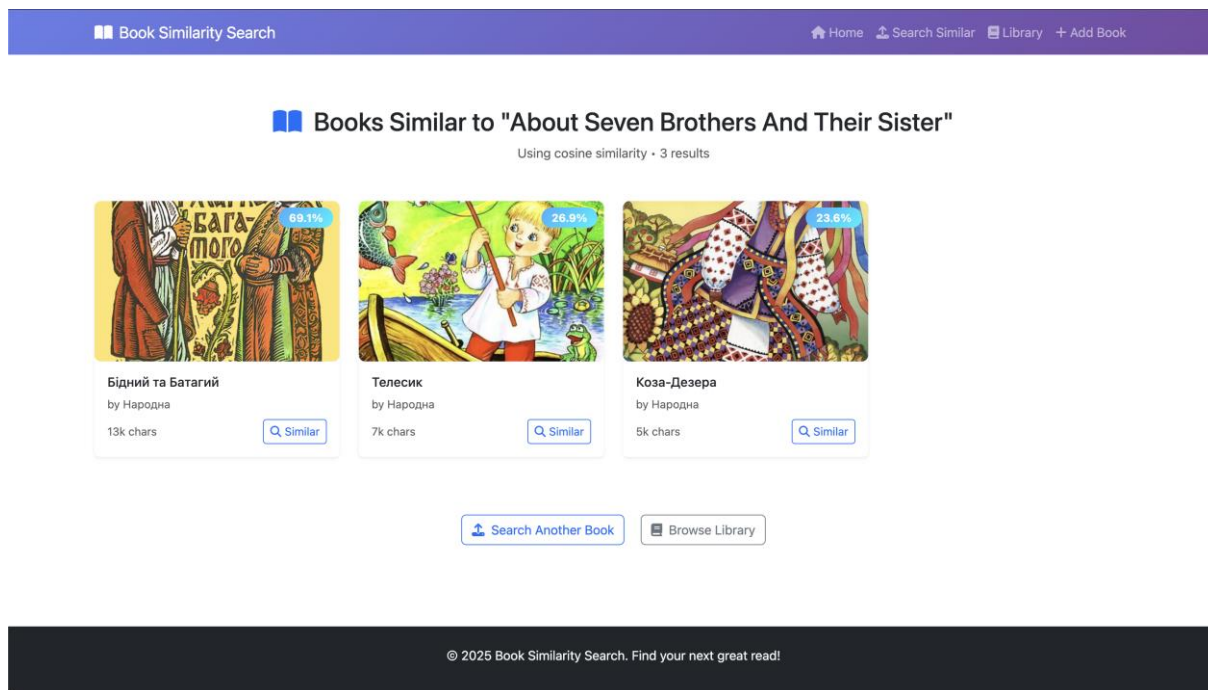


Рисунок 4.5 – Інтерфейс сторінки із рекомендованими казками

Кожна рекомендована книга відображається з відсотком схожості, який помітно виділений кольоровим значком у верхньому правому куті обкладинки книги. Цей кількісний показник схожості дозволяє користувачам миттєво зрозуміти силу рекомендації, зберігаючи при цьому візуальну привабливість.

Результати організовані в тому ж картковому макеті, що використовується в усій системі, що забезпечує узгодженість інтерфейсу. Кожна картка підтримує ті самі моделі взаємодії, що й інші елементи, включаючи ефекти наведення курсору та кнопки дій, які тепер містять як «Схожі» (для пошуку книг, схожих на цю рекомендацію), так і навігацію назад до головного інтерфейсу пошуку.

Заголовок сторінки контекстуалізує результати, повторюючи назву та автора книги, що шукається, а також метод схожості, що використовується, та кількість результатів. Ця інформація допомагає користувачам зрозуміти основу рекомендацій і дозволяє їм змінювати параметри для наступних пошуків.

У всій програмі кілька моделей взаємодії створюють цілісний користувацький досвід. Флеш-повідомлення надають системний відгук на дії користувача, з'являючись на всіх сторінках з можливістю автоматичного закриття. Стани завантаження реалізовані за допомогою анімації спінера та вимкнених кнопок, що створює відповідні очікування щодо часу обробки.

Система підтримує кілька точок входу та шляхів користувачів. Новачки можуть почати з завантаження однієї книги для отримання негайних рекомендацій, а досвідчені користувачі можуть почати з побудови комплексної бібліотеки для систематичного дослідження. Інтерфейс підтримує обидва підходи завдяки гнучкій навігації та контекстно-залежним опціям дій.

Обробка помилок реалізована за допомогою зручних повідомлень, які уникають технічних деталей, надаючи практичні рекомендації. Обмеження формату файлів, обмеження розміру та помилки обробки повідомляються простою мовою з пропозиціями щодо вирішення.

## ВИСНОВКИ

Ця кваліфікаційна робота успішно демонструє розробку та впровадження комплексної системи рекомендацій для казок та літературних творів, використовуючи сучасні методи машинного навчання та принципи дизайну, орієнтованого на користувача. Результатом аналітичного дослідження стало створення доступної платформи, яка дозволяє користувачам знаходити подібні книги за допомогою семантичного аналізу тексту, заповнюючи прогалину між передовими можливостями обробки природної мови та практичними потребами у пошуку літературних творів.

Впроваджена система демонструє кілька значних технічних досягнень. По-перше, розробка надійного конвеєра вилучення тексту, здатного обробляти різні формати файлів, включаючи PDF, EPUB, DOC, DOCX, FB2 та RTF, забезпечує широку сумісність з існуючими цифровими бібліотеками. Ця підтримка різних форматів вирішує критичну практичну проблему в реальному розгортанні, де користувачі мають різноманітні колекції, що охоплюють різні цифрові формати та джерела.

По-друге, інтеграція гнучкого генерування вбудовування за допомогою моделей Ollama та HuggingFace забезпечує масштабованість та адаптивність до різних обчислювальних середовищ. Цей підхід з двома моделями демонструє продуману архітектуру системи, яка враховує як сценарії локального впровадження, так і хмарні реалізації, дозволяючи при цьому вибирати модель на основі конкретних вимог до продуктивності та обмежень ресурсів.

Система рекомендацій демонструє складну технічну конструкцію завдяки впровадженню декількох методів обчислення схожості – косинусної схожості, евклідової відстані та скалярного добутку – що дозволяє користувачам досліджувати різні математичні підходи до схожості контенту. Така гнучкість забезпечує точне налаштування рекомендацій та

дає уявлення про те, як різні показники схожості працюють для різних типів літературного контенту.

Архітектура веб-додатку демонструє професійні практики розробки завдяки модульній конструкції, комплексному обробленню помилок та чуйному користувацькому інтерфейсу. Бекенд на базі Flask успішно управляє складними робочими процесами, що включають завантаження файлів, обробку тексту, генерацію вбудовування та обчислення схожості, зберігаючи при цьому стабільність системи та зручну для користувача систему повідомлення про помилки.

Особливо варто відзначити можливості системи з управління метаданими, яка автоматично витягує назви книг та імена авторів із завантаженого контенту, надаючи при цьому можливість ручного перезапису. Ця інтелектуальна автоматизація зменшує навантаження на користувача, зберігаючи при цьому гнучкість для крайніх випадків, коли автоматичне виявлення виявляється недостатнім.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Lessa W. A., Aarne A., Thompson S. The Types of the Folktale. A Classification and Bibliography. *The Journal of American Folklore*. 1963. Т. 76, № 301. С. 265. URL: <https://doi.org/10.2307/538541> (дата звернення: 14.05.2025).
2. Bacchilega C. *Postmodern Fairy Tales: Gender and Narrative Strategies*. University of Pennsylvania Press, 2010. 208 с.
3. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches / К. Cho та ін. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, м. Doha, Qatar. Stroudsburg, PA, USA, 2014. URL: <https://doi.org/10.3115/v1/w14-4012> (дата звернення: 14.05.2025).
4. Aftan S., Shah H. A Survey on BERT and Its Applications. *2023 20th Learning and Technology Conference (L&T)*, м. Jeddah, Saudi Arabia, 26 січ. 2023 р. 2023. URL: <https://doi.org/10.1109/lt58159.2023.10092289> (дата звернення: 14.05.2025).
5. Recommender systems survey / J. Bobadilla та ін. *Knowledge-Based Systems*. 2013. Т. 46. С. 109–132. URL: <https://doi.org/10.1016/j.knosys.2013.03.012> (дата звернення: 14.05.2025).
6. Research-paper recommender systems: a literature survey / J. Beel та ін. *International Journal on Digital Libraries*. 2015. Т. 17, № 4. С. 305–338. URL: <https://doi.org/10.1007/s00799-015-0156-0> (дата звернення: 14.05.2025).
7. Egger R. Text Representations and Word Embeddings. *Applied Data Science in Tourism*. Cham, 2022. С. 335–361. URL: [https://doi.org/10.1007/978-3-030-88389-8\\_16](https://doi.org/10.1007/978-3-030-88389-8_16) (дата звернення: 14.05.2025).
8. Statistical Significance in NLP / R. Dror та ін. *Statistical Significance Testing for Natural Language Processing*. Cham, 2020. С. 23–33. URL: [https://doi.org/10.1007/978-3-031-02174-9\\_4](https://doi.org/10.1007/978-3-031-02174-9_4) (дата звернення: 14.05.2025).

9. Gao T., Fisch A., Chen D. Making Pre-trained Language Models Better Few-shot Learners. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, м. Online. Stroudsburg, PA, USA, 2021. URL: <https://doi.org/10.18653/v1/2021.acl-long.295> (дата звернення: 14.05.2025).

10. Wells K. Analysis of Narrative Structure. *Narrative Inquiry*. 2011. С. 62–81. URL: <https://doi.org/10.1093/acprof:oso/9780195385793.003.0005> (дата звернення: 14.05.2025).

11. Mercy Milcah Y, Moorthi K. AI based Book Recommender System with Hybrid Approach. *International Journal of Engineering Research and*. 2020. Т. V9, № 02. URL: <https://doi.org/10.17577/ijertv9is020416> (дата звернення: 13.06.2025).

12. Pattern-based hybrid book recommendation system using semantic relationships / F. Wayesa та ін. *Scientific Reports*. 2023. Т. 13, № 1. URL: <https://doi.org/10.1038/s41598-023-30987-0> (дата звернення: 18.06.2025).

13. Personalized Book Recommendations: A Hybrid Approach Leveraging Collaborative Filtering, Association Rule Mining, and Content-Based Filtering / A. Bhajantri та ін. *EAI Endorsed Transactions on Internet of Things*. 2024. Т. 10. URL: <https://doi.org/10.4108/eetiot.6996> (дата звернення: 18.06.2025).

14. Lavrač N., Podpečan V., Robnik-Šikonja M. Text Embeddings. *Representation Learning*. Cham, 2021. С. 55–82. URL: [https://doi.org/10.1007/978-3-030-68817-2\\_3](https://doi.org/10.1007/978-3-030-68817-2_3) (дата звернення: 13.06.2025).

15. Dmytrenko V. Artistic exploitation of creativity in ukrainian social fairy tales. *International Humanitarian University Herald. Philology*. 2020. Т. 3, № 46. С. 57–60. URL: <https://doi.org/10.32841/2409-1154.2020.46-3.13> (дата звернення: 13.06.2025).

16. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing / P. Liu та ін. *ACM Computing Surveys*. 2022. URL: <https://doi.org/10.1145/3560815> (дата звернення: 13.06.2025).

17. Bendix R., Luthi M., Niles J. D. The European Folktale: Form and Nature. *Western Folklore*. 1983. Т. 42, № 4. С. 311. URL: <https://doi.org/10.2307/1499511> (дата звернення: 13.06.2025).

18. Mohammad S. M. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*. 2012. Т. 53, № 4. С. 730–741. URL: <https://doi.org/10.1016/j.dss.2012.05.030> (дата звернення: 13.06.2025).

19. World literature for children. *Children's Literature Comes of Age*. 2015. С. 27–60. URL: <https://doi.org/10.4324/9781315667492-10> (дата звернення: 13.06.2025).

20. Morphology of the Folktale / M. Jacobs та ін. *The Journal of American Folklore*. 1959. Т. 72, № 284. С. 195. URL: <https://doi.org/10.2307/538485> (дата звернення: 13.06.2025).